(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 December 2001 (06.12.2001)

PCT

(10) International Publication Number
WO 01/92495 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US01/17405

(22) International Filing Date: 29 May 2001 (29.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/207,650 26 May 2000 (26.05.2000) US

(71) Applicant *(for all designated States except US)*: **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **MEYERS, Rachel** [US/US]; 115 Devonshire Road, Newton, MA 02468 (US).

(74) Agents: **MANDRAGOURAS, Amy, E.**; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/92495 A2

(54) **Title:** 47765, A NOVEL HUMAN LYSYL OXIDASE AND USES THEREOF

(57) **Abstract:** The invention provides isolated nucleic acid molecules, designated LSO nucleic acid molecules, which encode novel LSO-related lysyl oxidase molecules. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing LSO nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which an LSO gene has been introduced or disrupted. The invention still further provides isolated LSO proteins, fusion proteins, antigenic peptides and anti-LSO antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

47765, A NOVEL HUMAN LYSYL OXIDASE AND USES THEREOF

Related Applications

This application claims priority to U.S. Provisional Application Serial No.: 60/207,650 filed on May 26, 2000, incorporated herein in its entirety by this reference.

Background of the Invention

Biogenesis of connective tissue matrices is a prerequisite for the production and functioning of a number of human tissues, including skin, muscle, and bone. A family of enzymes that facilitates the covalent cross-linking of the molecular units of collagen and elastin, termed the lysyl oxidase family, has been identified. These enzymes catalyze the oxidation of peptidyl lysine in these extracellular matrix proteins to peptidyl α-aminoadipic-δ-semialdehyde, which is able to condense with neighboring amino groups or peptidyl aldehydes to form covalent crosslinkages (Kagan (1986) in: Biology of the Extracellular Matrix, Mecham, ed. Vol. I: Regulation of Matrix Accumulation. Acadmeic Press: Orlando, FL: 321-398).

Members of the lysyl oxidase family have been isolated from a wide variety of animal species, including human, rodent, avian, mammalian, and piscine sources. A protein having activity and structure similar to that of lysyl oxidase has also been isolated from yeast (Dove *et al.*(1996) *FEBS Lett.* 398: 231-234). Family members vary widely in structure, having between 48 and 100% homology (Smith-Mungo and Kagan (1998) *Matrix Biology* 16: 387-398). The region of the protein which has been found to share the greatest identity (90-95%) between family members is the C-terminal segment of the preproprotein amino acid sequence (Smith-Mungo and Kagan (1998), *supra*). This region is believed to include the active site, as well as sequences involved in the octahedral coordination of the copper ion (Gacheru *et al.* (1990) *J. Biol. Chem.* 265: 19022-19027), and a binding site for another cofactor, lysyltyrosine quinone (Wang et al. (1996). Lysyl oxidases are also characterized by an N-terminal signal peptide (70-72% identical), which is thought to mediate the secretion of these proteins into the surrounding extracellular matrix (Smith-Mungo and Kagan (1998), *supra*).

Studies of lysyl oxidase have indicated that this enzyme undergoes significant trafficking and processing prior to becoming fully functional in the cell. For example, bovine lysyl oxidase is synthesized as a 46 kDa sequence having an N-terminal 21 residue signal peptide. The N-terminal portion of the protein is N-glycosylated at two or three sites and undergoes signal peptide cleavage to yield an inactive 50 kD proenzyme,

which is secreted into the extracellular matrix (Trackman *et al.* (1990) *Biochemistry* 29: 4863-4870; Trackman *et al.* (1991) *Biochemistry* 30: 8282). The proenzyme is cleaved at a Gly-Asp bond by a metalloproteinase (also present in the extracellular matrix) to the fully-active 32 kD species, lacking the N-glycosylated N-terminal region (Trackman *et al.* (1992) *J. Biol. Chem.* 267: 8666-8671). It is thought that this process prevents lysyl oxidase from catalyzing the cross-linking of nascent matrix macromolecules prior to their export from the cell (Smith-Mungo and Kagan (1998), *supra*).

The mechanism by which lysyl oxidases catalyze the covalent cross-linking of the molecular subunits of elastin and collagen has also been characterized. Lysyl oxidase first forms a Schiff base with its carbonyl cofactor (lysyltyrosine quinone in the case of animal lysyl oxidase; trihydroxyphenylalanine quinone in the case of yeast lysyl oxidase (Dove et al. (1996), *supra*). The rate-limiting step is the general base (histidine residue)-facilitated α-proton abstraction from the substrate (Gacheru *et al.* (1988) *J. Biol. Chem.* 265: 19022-19027; Williamson and Kagan (1987) *J. Biol. Chem.* 261: 9477-9482). The carbonyl cofactor is reduced by electron migration from the substrate carbanion, and the aldehyde product is released. The reduced enzyme, still bound to the amino group of the substrate, is reoxidized by molecular oxygen to produce hydrogen peroxide and ammonia (Smith-Mungo and Kagan (1998), *supra*). While elastin and collagen subunits are typically the substrates for lysyl oxidase, *in vitro* studies have demonstrated that this enzyme is able to oxidize peptidyl lysine in a number of different basic, globular proteins (Kagan *et al.* (1984) *J. Biol. Chem.* 259: 11203-11207). A coordinated copper ion mediates the activity of the enzyme in this reaction mechanism.

Lysyl oxidases play an important role in the production of connective matrices, such as elastin or collagen matrices. Such matrices are of vital importance in maintaining the structure of the cell (*e.g.*, the extracellular matrix, or cell wall). As such, their activity contributes to the ability of the cell to grow, differentiate and proliferate. On a larger scale, such matrices are also critical for the formation of various tissues, such as skin, muscle, bone, and cartilage. Underscoring the importance of this family of enzymes, modulation of the activity of one or more lysyl oxidases has been linked to a number of human diseases, including cutis laxa and Ehlers-Danlos syndrome type V (both skin elasticity disorders) (Khakoo *et al.* (1997) *Clin. Genet.* 51: 109-114; DiFerrante *et al.* (1975) *Connect. Tissue Res.* 3: 49-53), and Menkes' disease (Kuivaniemi *et al.* (1985) *Am. J. Hum. Genet.* 37: 798-808). The human lysyl oxidase

gene has been mapped to human chromosome 5q23.3-31.2 (Hamalainen *et al.* (1991) *Genomics* 11: 508-516; Svinarich *et al.* (1992) *J. Biol. Chem.* 267: 14382-14387; and Mariani *et al.* (1992)) *Matrix* 12: 242-248), and disorders linked to this region may also involve lysyl oxidase activity.

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel members of the family of lysyl oxidase molecules, referred to herein as LSO nucleic acid and protein molecules. The LSO nucleic acid and protein molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, *e.g.*, cellular proliferation, growth, differentiation, or migration. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding LSO proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of LSO-encoding nucleic acids.

In one embodiment, an LSO nucleic acid molecule of the invention is at least 42%, 48%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the nucleotide sequence (*e.g.*, to the entire length of the nucleotide sequence) shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a complement thereof.

In a preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown in SEQ ID NO:1 or 3, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:3 and nucleotides 1-94 of SEQ ID NO:1. In yet a further embodiment, the nucleic acid molecule includes SEQ ID NO:3 and nucleotides 2366-2976 of SEQ ID NO:1. In another preferred embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:1 or 3, or a complement thereof. In another preferred embodiment, the nucleic acid molecule includes a fragment of at least 1243 or1412 nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, or a complement thereof.

In preferred embodiments, the nucleic acid molecules are at least 15 (*e.g.*, contiguous) nucleotides in length and hybridize under stringent conditions to nucleotides 1-32 and 556-571 of SEQ ID NO:1. In other preferred embodiments, the nucleic acid molecules comprise nucleotides 1-32 and 556-572 of SEQ ID NO:1. In another

preferred embodiment, the nucleic acid molecules consist of nucleotides 1-32 and 556-572 of SEQ ID NO:1.

In another embodiment, an LSO nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In a preferred embodiment, an LSO nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 42%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the entire length of the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

In another preferred embodiment, an isolated nucleic acid molecule encodes the amino acid sequence of human LSO. In yet another preferred embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In yet another preferred embodiment, the nucleic acid molecule is at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1242, 1250, 1300, 1350, 1400, 1412, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900 or more nucleotides in length. In a further preferred embodiment, the nucleic acid molecule is at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1242, 1250, 1300, 1350, 1400, 1412, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900 or more nucleotides in length and encodes a protein having an LSO activity (as described herein).

Another embodiment of the invention features nucleic acid molecules, preferably LSO nucleic acid molecules, which specifically detect LSO nucleic acid molecules relative to nucleic acid molecules encoding non-LSO proteins. For example, in one embodiment, such a nucleic acid molecule is at least 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800,

1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900 or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1 or 3, or a complement thereof, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

In preferred embodiments, the nucleic acid molecules are at least 15 (*e.g.*, 15 contiguous) nucleotides in length and hybridize under stringent conditions to the nucleotide molecule set forth in SEQ ID NO:1, or a complement thereof.

In other preferred embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____, wherein the nucleic acid molecule hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1 or 3, respectively, under stringent conditions.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to an LSO nucleic acid molecule, *e.g.*, the coding strand of an LSO nucleic acid molecule.

Another aspect of the invention provides a vector comprising an LSO nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment, the invention provides a host cell containing a vector of the invention. In yet another embodiment, the invention provides a host cell containing a nucleic acid molecule of the invention. The invention also provides a method for producing a protein, preferably an LSO protein, by culturing in a suitable medium, a host cell, *e.g.*, a mammalian host cell such as a non-human mammalian cell, of the invention containing a recombinant expression vector, such that the protein is produced.

Another aspect of this invention features isolated or recombinant LSO proteins and polypeptides. In one embodiment, an isolated LSO protein includes at least one or more of the following motifs or domains: a signal peptide, an N-glycosylation site, a lysyl oxidase domain, an LSO signature motif, and/or a scavenger receptor cysteine-rich domain.

In a preferred embodiment, an LSO protein includes at least one or more of the following motifs or domains: a signal peptide, an N-glycosylation site, a lysyl oxidase domain, an LSO signature motif, and/or a scavenger receptor cysteine-rich domain, and

has an amino acid sequence at least about 42%, 50%, 55%, 60%, 65%, 67%, 68%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

In another preferred embodiment, an LSO protein includes at least one or more of the following motifs or domains: a signal peptide, an N-glycosylation site, a lysyl oxidase domain, an LSO signature motif, and/or a scavenger receptor cysteine-rich domain, and has an LSO activity (as described herein).

In yet another preferred embodiment, an LSO protein includes at least one or more of the following motifs or domains: a signal peptide, an N-glycosylation site, a lysyl oxidase domain, an LSO signature motif, and/or a scavenger receptor cysteine-rich domain, and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3.

In another embodiment, the invention features fragments of the protein having the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 412 amino acids (*e.g.*, contiguous amino acids) of the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____. In another embodiment, an LSO protein has the amino acid sequence of SEQ ID NO:2. In yet another embodiment, an LSO protein comprises or consists of amino acid residues 1-31 and/or 85-90 of SEQ ID NO:2.

In another embodiment, the invention features an LSO protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 42%, 48%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to a nucleotide sequence of SEQ ID NO:1 or 3, or a complement thereof. This invention further features an LSO protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3, or a complement thereof.

The proteins of the present invention or portions thereof, *e.g.*, biologically active portions thereof, can be operatively linked to a non-LSO polypeptide (*e.g.*, heterologous amino acid sequences) to form fusion proteins. The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind proteins

of the invention, preferably LSO proteins. In addition, the LSO proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of an LSO nucleic acid molecule, protein, or polypeptide in a biological sample by contacting the biological sample with an agent capable of detecting an LSO nucleic acid molecule, protein, or polypeptide such that the presence of an LSO nucleic acid molecule, protein or polypeptide is detected in the biological sample.

In another aspect, the present invention provides a method for detecting the presence of LSO activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of LSO activity such that the presence of LSO activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating LSO activity comprising contacting a cell capable of expressing LSO with an agent that modulates LSO activity such that LSO activity in the cell is modulated. In one embodiment, the agent inhibits LSO activity. In another embodiment, the agent stimulates LSO activity. In one embodiment, the agent is an antibody that specifically binds to an LSO protein. In another embodiment, the agent modulates expression of LSO by modulating transcription of an LSO gene or translation of an LSO mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an LSO mRNA or an LSO gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant or unwanted LSO protein or nucleic acid expression or activity by administering an agent which is an LSO modulator to the subject. In one embodiment, the LSO modulator is an LSO protein. In another embodiment the LSO modulator is an LSO nucleic acid molecule. In yet another embodiment, the LSO modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant or unwanted LSO protein or nucleic acid expression is a lysyl oxidase-associated disorder, *e.g.*, a cell proliferation, growth, or differentiation disorder, a muscular disorder, a bone disorder, a skin elasticity disorder, or a cartilage-based disorder.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding an LSO protein; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of an LSO protein, wherein a wild-type form of the gene encodes a protein with an LSO activity.

In another aspect the invention provides methods for identifying a compound that binds to or modulates the activity of an LSO protein, by providing an indicator composition comprising an LSO protein having LSO activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on LSO activity in the indicator composition to identify a compound that modulates the activity of an LSO protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence and predicted amino acid sequence of human LSO (clone Fbh47765). The nucleotide sequence corresponds to nucleic acids 1 to 2976 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 756 of SEQ ID NO:2. The coding region without the 3' untranslated region of the human LSO gene is shown in SEQ ID NO:3.

Figure 2 depicts a structural, hydrophobicity, and antigenicity analysis of the human LSO protein (SEQ ID NO:2).

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to interchangeably herein as "lysyl oxidase" or "LSO" nucleic acid and protein molecules (e.g., human LSO or LSO47765), which are novel members of a family of enzymes possessing lysyl oxidase activity.

As used herein, the term "lysyl oxidase" includes a molecule which is involved in the oxidation of peptidyl lysine to peptidyl α-aminoadipic-δ semialdehyde, thereby catalyzing the covalent crosslinking between and within the molecular units of extracellular matrix proteins, *e.g.,* elastin or collagen. Lysyl oxidase molecules are involved in the biogenesis of connective tissue matrices, and therefore are involved in cellular growth, proliferation, and differentiation. Lysyl oxidase molecules are also involved in systemic processes such as skin, bone, cartilage and muscle formation and function, and in tumor formation. Thus, the LSO molecules of the present invention provide novel diagnostic targets and therapeutic agents to control lysyl oxidase-associated disorders.

As used herein, a "lysyl oxidase-associated disorder" includes a disorder, disease or condition which is caused or characterized by a misregulation (*e.g.*, downregulation or upregulation) of lysyl oxidase activity. Lysyl oxidase-associated disorders can detrimentally affect cellular functions such as cellular proliferation, growth, differentiation, or migration; or systemic responses in an organism, such as bone, skin, cartilage, or muscle formation, structure, or elasticity, or tumor formation. Examples of lysyl oxidase-associated disorders include muscular disorders, such as cardiac muscle-related disorders. Cardiovascular system disorders in which the LSO molecules of the invention may be directly or indirectly involved include arteriosclerosis, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, atrial fibrilation, Jervell syndrome, Lange-Nielsen syndrome, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, and arrhythmia. LSO-mediated or related disorders also include disorders of the musculoskeletal system such as paralysis and muscle weakness, *e.g.*, ataxia, myotonia, and myokymia.

Lysyl oxidase disorders also include cellular proliferation, growth, differentiation, or migration disorders. Cellular proliferation, growth, differentiation, or migration disorders include those disorders that affect cell proliferation, growth, differentiation, or migration processes. As used herein, a "cellular proliferation, growth,

differentiation, or migration process" is a process by which a cell increases in number, size or content, by which a cell develops a specialized set of characteristics which differ from that of other cells, or by which a cell moves closer to or further from a particular location or stimulus. The LSO molecules of the invention are known to be involved in production/maintenance of the extracellular matrix, which is of vital importance in maintaining the structure of the cell. Thus, the LSO molecules may modulate cellular growth, differentiation, or migration, and may play a role in disorders characterized by aberrantly regulated growth, differentiation, or migration. Such disorders include cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; hepatic disorders; and hematopoietic and/or myeloproliferative disorders.

LSO-associated or related disorders also include disorders of bone formation and resorption, including osteoporosis, osteochondrosis, and osteopetrosis.

LSO-associated or related disorders also include disorders of skin formation and elasticity, including cutis laxa and Ehlers-Danlos type V syndrome.

LSO-associated or related disorders also include disorders of cartilage formation and structure, including chondromalacia and polychondritis.

LSO-associated or related disorders also include disorders affecting tissues in which LSO protein is expressed.

As used herein, a "lysyl oxidase-mediated activity" includes an activity mediated by a lysyl oxidase polypeptide. A lysyl oxidase-mediated activity includes the catalysis of the oxidation of peptidyl lysine to peptidyl α-aminoadipic-δ semialdehyde, thereby catalyzing the covalent crosslinking between and within the molecular units of extracellular matrix proteins, *e.g.*, elastin and of collagen. Lysyl oxidase-mediated activities include those cellular or systemic activities which require the cross-linking of extracellular matrix proteins, *e.g.*, collagen or elastin. Such activities include cellular growth, proliferation, and differentation, and also systemic activities, such as bone, skin, cartilage and muscle biosynthesis and function, or tumor formation.

The term "family" when referring to the protein and nucleic acid molecules of the invention (*e.g.*, the LSO family of proteins and/or nucleic acids) is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be

from either the same or different species. For example, a family can contain a first protein of human origin, as well as other, distinct proteins of human origin or alternatively, can contain homologues of non-human origin, *e.g.*, mouse or monkey proteins. Members of a family may also have common functional characteristics.

For example, proteins that belong to the family of LSO proteins comprise at least one signal sequence or signal peptide. The prediction of such a signal peptide can be made, for example, utilizing the computer algorithm SignalP (Henrik, *et al.* (1997) *Protein Engineering* 10:1-6). As used herein, a "signal sequence" or "signal peptide" includes a peptide containing about 15 or more amino acids which occurs at the N-terminus of secretory and membrane bound proteins and which contains a large number of hydrophobic amino acid residues. For example, a signal sequence contains at least about 10-30 amino acid residues, preferably about 15-25 amino acid residues, more preferably about 18-20 amino acid residues, and more preferably about 19 amino acid residues, and has at least about 35-65%, preferably about 38-50%, and more preferably about 40-45% hydrophobic amino acid residues (*e.g.*, Valine, Leucine, Isoleucine or Phenylalanine). Such a "signal sequence", also referred to in the art as a "signal peptide", serves to direct a protein containing such a sequence to a lipid bilayer, and is cleaved in secreted and membrane bound proteins. A signal sequence was identified in the amino acid sequence of human LSO at about amino acids 1-24 of SEQ ID NO:2.

In another embodiment, an LSO molecule of the present invention is identified based on the presence of at least one N-glycosylation site. As used herein, the term "N-glycosylation site" includes an amino acid sequence of about 4 amino acid residues in length which serves as a glycosylation site. More preferably, an N-glycosylation site has the consensus sequence Asn-Xaa-Ser/Thr (where Xaa may be any amino acid). N-glycosylation sites are described in, for example, Prosite PDOC00001 (http://www.expasy.ch/cgi-bin/get-prodoc-entry?PDOC00001), the contents of which are incorporated herein by reference. Amino acid residues 198-201 and 629-632 of the LSO protein comprise N-glycosylation sites. Accordingly, LSO proteins having at least one N-glycosylation site are within the scope of the invention.

In another embodiment, an LSO molecule of the present invention is identified based on the presence of a "lysyl oxidase domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "lysyl oxidase domain" includes a protein domain having an amino acid sequence of about 100-300 amino acid residues,

and a bit score of at least 200 when compared against a lysyl oxidase Hidden Markov Model (HMM), e.g., PFAM accession number PF01186. In a preferred embodiment, a lysyl oxidase domain includes a protein domain having an amino acid sequence of about 150-250 amino acid residues and a bit score of at least 300. In another preferred embodiment, a lysyl oxidase domain includes a protein domain having an amino acid sequence of about 202-205 amino acid residues and a bit score of at least 475 (*e.g.*, 480, 490, 500, 510, 513 or more). To identify the presence of a lysyl oxidase domain in an LSO protein, the amino acid sequence of the protein is used to search a database of known Hidden Markov Models (HMMs e.g., the PFAM HMM database). The lysyl oxidase domain (HMM) has been assigned the PFAM Accession PF01186 (http://genome.wustl.edu/Pfam/html). For example, a search was performed against the HMM database using the amino acid sequence (SEQ ID NO:2) of human LSO, resulting in the identification of a lysyl oxidase domain in the amino acid sequence of human LSO (SEQ ID NO:2) at about residues 533-736 of SEQ ID NO:2, having a score of 513.1.

In another embodiment of the invention, an LSO protein is identified based on the presence of at least one "LSO signature motif" in the protein or corresponding nucleic acid molecule. As used herein, the term "LSO signature motif" includes an amino acid sequence that contains at least about 5-20 amino acid residues that are conserved among LSO family members. In one embodiment, an LSO signature motif includes an amino acid sequence at least about 7-17 amino acid residues, more preferably about 9-15 amino acid residues, more preferably 10-13 amino acid residues and still more preferably 11 amino acid residues in length and having the following amino acid sequence: W-X-W-H-X-C-H-X-H-Y-H, (SEQ ID NO:4), where X indicates any amino acid (see, for example, Krebs and Krawetz (1993) *Biochim. Biophys. Acta* 1202: 7-12). Accordingly, preferred proteins include the conserved amino acid residues of the above-recited LSO signature motif. Proteins including at least 7, 8, 9, 10 or more conserved amino acid residues of the above-recited LSO signature motif are also considered to be within the scope of the present invention.

In another embodiment, an LSO molecule of the present invention is identified based on the presence of a "scavenger receptor cysteine-rich domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "scavenger receptor cysteine-rich domain" includes a protein domain having an amino acid sequence of

about 50-150 amino acid residues, and a bit score of at least 25 when compared against a scavenger receptor cysteine-rich domain Hidden Markov Model (HMM), *e.g.*, PFAM accession number PF00530. In a preferred embodiment, a scavenger receptor cysteine-rich domain includes a protein domain having an amino acid sequence of about 75-125 amino acid residues and a bit score of at least 35. In another preferred embodiment, a scavenger receptor cysteine-rich domain includes a protein domain having an amino acid sequence of about 97-119 amino acid residues and a bit score of at least 40 (*e.g.*, 41, 42, 43, 44, 45, 46 or higher). To identify the presence of a scavenger receptor cysteine-rich domain in an LSO protein, the amino acid sequence of the protein may be used to search a database of known Hidden Markov Models (HMMs *e.g.*, the PFAM HMM database). The scavenger receptor cysteine-rich domain (HMM) has been assigned the PFAM Accession PF00530 (http://genome.wustl.edu/Pfam/html). For example, a search was performed against the HMM database using the amino acid sequence (SEQ ID NO:2) of human LSO, resulting in the identification of four scavenger receptor cysteine-rich domains in the amino acid sequence of human LSO (SEQ ID NO:2) at about residues 37-133, 169-287, 314-411, and 424-529 of SEQ ID NO:2, having scores of 98.1, 30.4, 115.8, and 46.3, respectively.

In a preferred embodiment, the LSO molecules of the invention include at least one, preferably two or more of the following motifs or domains: a signal peptide, an N-glycosylation site, a lysyl oxidase domain, an LSO signature motif, and/or a scavenger receptor cysteine-rich domain.

Isolated proteins of the present invention, preferably LSO proteins, have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:2, or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NO:1 or 3. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least 30%, 40%, or 50% homology, preferably 60% homology, more preferably 70%-80%, and even more preferably 90-95% homology across the amino acid sequences of the domains and contain at least one and

preferably two structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide sequences which share at least 30%, 40%, or 50%, preferably 60%, more preferably 70-80%, or 90-95% homology and share a common functional activity are defined herein as sufficiently identical.

As used interchangeably herein, an "LSO activity", "biological activity of LSO" or "functional activity of LSO", refers to an activity exerted by an LSO protein, polypeptide or nucleic acid molecule on an LSO responsive cell or tissue, or on an LSO protein substrate, as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, an LSO activity is a direct activity, such as an association with an LSO-target molecule. As used herein, a "target molecule" or "binding partner" is a molecule with which an LSO protein binds or interacts in nature, such that LSO-mediated function is achieved. An LSO target molecule can be a non-LSO molecule (*e.g.*, a copper ion, lysyltyrosine quinone, trihydroxyphenylalanine quinone, or other cofactor) or an LSO protein or polypeptide of the present invention. In an exemplary embodiment, an LSO target molecule is an LSO substrate (*e.g.*, elastin or collagen). Alternatively, an LSO activity is an indirect activity, such as a metabolic activity mediated by interaction of the LSO protein with an LSO substrate. The biological activities of LSO are described herein. In an exemplary embodiment, the LSO proteins of the present invention have at least one of the following activities: i) interaction with an LSO substrate; ii) interaction with a cofactor (*e.g.*, a copper ion, lysyltyrosine quinone, trihydroxyphenylalanine quinone, or other cofactor); and iii) conversion of an LSO substrate to product (*e.g.*, catalysis of the conversion of substrate to product). In yet another embodiment, the LSO proteins of the present invention have one or more of the following activities: 1) modulate cellular growth, proliferation, or differentiation, 2) modulate skin formation or elasticity 3) modulate bone formation or structure; 4) modulate muscle formation or elasticity; 5) modulate cartilage formation or structure; and 6) modulate tumor formation.

Accordingly, another embodiment of the invention features isolated LSO proteins and polypeptides having an LSO activity. Other preferred proteins are LSO proteins having one or more of the following motifs or domains: a signal peptide, an N-glycosylation site, a lysyl oxidase domain, an LSO signature motif, and/or a scavenger receptor cysteine-rich domain and, preferably, an LSO activity.

Additional preferred proteins have at least one or more of the following motifs or domains: a signal peptide, an N-glycosylation site, a lysyl oxidase domain, an LSO signature motif, and/or a scavenger receptor cysteine-rich domain, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3.

The nucleotide sequence of the isolated human LSO cDNA and the predicted amino acid sequence of the human LSO polypeptide are shown in Figure 1 and in SEQ ID NOs:1 and 2, respectively. A plasmid containing the nucleotide sequence encoding human LSO was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Numbers _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human LSO gene, which is approximately 2976 nucleotides in length, encodes a protein having a molecular weight of approximately 83.2 kD and which is approximately 756 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode LSO proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify LSO-encoding nucleic acid molecules (*e.g.*, LSO mRNA) and fragments for use as PCR primers for the amplification or mutation of LSO nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated LSO nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ as a hybridization probe, LSO nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

A nucleic acid of the invention can be amplified using cDNA, mRNA or, alternatively, genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to LSO nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1 or 3. This cDNA may comprise sequences encoding the human LSO protein (*i.e.*, "the coding region", from nucleotides 95-2365), as well as 5' untranslated sequences (nucleotides 1-94) and 3' untranslated sequences (nucleotides 2366-2976) of SEQ ID NO:1. Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (*e.g.*, nucleotides 95-2365, corresponding to SEQ ID NO:3).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, respectively, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 42%, 48%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the entire length of the nucleotide sequence shown in SEQ ID NO:1 or 3, or the entire length of the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of an LSO protein, *e.g.*, a biologically active portion of an LSO protein. The nucleotide sequence determined from the cloning of the LSO gene allows for the generation of probes and primers designed for use in identifying and/or cloning other LSO family members, as well as LSO homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ of an anti-sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In one embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is greater than 20-50, 50-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1050, 1050-1100, 1100-1150, 1150-1200, 1200-1250, 1250-1300, 1300-1350, 1350-1400, 1400-1450, 1450-1500, 1500-1550, 1550-1600, 1600-1650, 1650-1700, 1700-1750, 1750-1800, 1800-1850, 1850-1900, 1900-1950, 1950-2000, 2000-2050, 2050-2100, 2100-2150, 2150-2200, 2200-2250, 2250-2300, 2300-2350, 2350-2400, 2400-2450, 2450-2500, 2500-2550, 2550-2600, 2600-2650, 2650-2700, 2700-2750, 2750-2800, 2800-2850, 2850-2900, 2900-2950 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

Ranges intermediate to the above-recited values, *e.g.*, nucleic acid molecules comprising the nucleic acid sequence which is 50-60, 60-70, 70-80, 80-90, 90-100 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ are also intended to be encompassed by the present invention. Values and ranges included and/or intermediate within the ranges set forth herein are also intended to be within the scope of the present invention. For example, nucleic acid molecules comprising the nucleic acid sequence which is 51, 52, 53, 54, 55, 56, 57, 58, and 59 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ are intended to be included within the range of 50-60 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

Probes based on the LSO nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress an LSO protein, such as by measuring a level of an LSO-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting LSO mRNA levels or determining whether a genomic LSO gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of an LSO protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ which encodes a polypeptide having an LSO biological activity (the biological activities of the LSO proteins are described herein), expressing the encoded portion of the LSO protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the LSO protein.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ due to degeneracy of the genetic code and thus encode the same LSO proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2.

In addition to the LSO nucleotide sequences shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the LSO proteins may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the LSO genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding an LSO protein, preferably a mammalian LSO protein, and can further include non-coding regulatory sequences, and introns.

Allelic variants of human LSO include both functional and non-functional LSO proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the human LSO protein that maintain the ability, for example, to bind an LSO ligand or substrate and/or to modulate cell growth, proliferation and/or differentiation mechanisms. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the human LSO protein that do not have the ability to either bind an LSO ligand and/or modulate any of the LSO activities described herein. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion or deletion in critical residues or critical regions of the protein.

The present invention further provides non-human orthologues of the human LSO protein. Orthologues of the human LSO protein are proteins that are isolated from non-human organisms and possess the same LSO ligand binding and/or modulation of membrane excitability activities of the human LSO protein. Orthologues of the human LSO protein can readily be identified as comprising an amino acid sequence that is substantially identical to SEQ ID NO:2.

Moreover, nucleic acid molecules encoding other LSO family members and, thus, which have a nucleotide sequence which differs from the LSO sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ are intended to be within the scope of the invention. For example, another LSO cDNA can be identified based on the nucleotide sequence of human LSO. Moreover, nucleic acid molecules encoding LSO proteins from different species, and which, thus, have a nucleotide sequence which differs from the LSO sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ are intended to be within the scope of the invention. For example, a mouse LSO cDNA can be identified based on the nucleotide sequence of a human LSO.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the LSO cDNAs of the invention can be isolated based on their homology to the LSO nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural allelic variants and homologues of the LSO cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the LSO gene.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to a complement of the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In other embodiment, the nucleic acid is at least 20-50, 50-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1050, 1050-1100, 1100-1150, 1150-1200, 1200-1250, 1250-1300, 1300-1350, 1350-1400, 1400-1450, 1450-

1500, 1500-1550, 1550-1600, 1600-1650, 1650-1700, 1700-1750, 1750-1800, 1800-1850, 1850-1900, 1900-1950, 1950-2000, 2000-2050, 2050-2100, 2100-2150, 2150-2200, 2200-2250, 2250-2300, 2300-2350, 2350-2400, 2400-2450, 2450-2500, 2500-2550, 2550-2600, 2600-2650, 2650-2700, 2700-2750, 2750-2800, 2800-2850, 2850-2900, 2900-2950 or more nucleotides in length.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences that are significantly identical or homologous to each other remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995), sections 2, 4 and 6. Additional stringent conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9 and 11. A preferred, non-limiting example of stringent hybridization conditions includes hybridization in 4X or 6X sodium chloride/sodium citrate (SSC), at about 65-70°C (or hybridization in 4X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 1X SSC, at about 65-70°C. A further preferred, non-limiting example of stringent hybridization conditions includes hybridization at 6X SSC at 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. A preferred, non-limiting example of highly stringent hybridization conditions includes hybridization in 1X SSC, at about 65-70°C (or hybridization in 1X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 0.3X SSC, at about 65-70°C. A preferred, non-limiting example of reduced stringency hybridization conditions includes hybridization in 4X or 6X SSC, at about 50-60°C (or alternatively hybridization in 6X SSC plus 50% formamide at about 40-45°C) followed by one or more washes in 2X SSC, at about 50-60°C. Ranges intermediate to the above-recited values, *e.g.*, at 65-70°C or at 42-50°C are also intended to be encompassed by the present invention. SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes each after hybridization is complete. The hybridization temperature for hybrids anticipated to be

less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, T_m(°C) = 2(# of A + T bases) + 4(# of G + C bases). For hybrids between 18 and 49 base pairs in length, T_m(°C) = 81.5 + 16.6($\log_{10}[Na^+]$) + 0.41(%G+C) - (600/N), where N is the number of bases in the hybrid, and $[Na^+]$ is the concentration of sodium ions in the hybridization buffer ($[Na^+]$ for 1xSSC = 0.165 M). It will also be recognized by the skilled practitioner that additional reagents may be added to hybridization and/or wash buffers to decrease non-specific hybridization of nucleic acid molecules to membranes, for example, nitrocellulose or nylon membranes, including but not limited to blocking agents (*e.g.*, BSA or salmon or herring sperm carrier DNA), detergents (*e.g.*, SDS), chelating agents (*e.g.*, EDTA), Ficoll, PVP and the like. When using nylon membranes, in particular, an additional preferred, non-limiting example of stringent hybridization conditions is hybridization in 0.25-0.5M NaH_2PO_4, 7% SDS at about 65°C, followed by one or more washes at 0.02M NaH_2PO_4, 1% SDS at 65°C, see *e.g.*, Church and Gilbert (1984) *Proc. Natl. Acad. Sci. USA* 81:1991-1995, (or alternatively 0.2X SSC, 1% SDS).

Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1 or 3 and corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*i.e.*, encodes a natural protein).

In addition to naturally-occurring allelic variants of the LSO sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, thereby leading to changes in the amino acid sequence of the encoded LSO proteins, without altering the functional ability of the LSO proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of LSO (*e.g.*, the sequence of SEQ ID NO:2) without altering the biological activity, whereas an "essential" amino acid residue is required for biological

activity. For example, amino acid residues that are conserved among the LSO proteins of the present invention, *e.g.*, those present in a lysyl oxidase consensus sequence, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the LSO proteins of the present invention and other members of the LSO family are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding LSO proteins that contain changes in amino acid residues that are not essential for activity. Such LSO proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 42%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2.

An isolated nucleic acid molecule encoding an LSO protein identical to the protein of SEQ ID NO:2 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an LSO protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be

introduced randomly along all or part of an LSO coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for LSO biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant LSO protein can be assayed for the ability to modulate cellular growth, proliferation, or differentiation, to modulate bone, skin, cartilage, or muscle growth, structure, or elasticity, or to modulate tumor formation.

In addition to the nucleic acid molecules encoding LSO proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire LSO coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an LSO. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region of human LSO corresponds to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding LSO. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding LSO disclosed herein (*e.g.*, SEQ ID NO:3), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of LSO mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of LSO mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of LSO mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length.

An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an LSO protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified

to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave LSO mRNA transcripts to thereby inhibit translation of LSO mRNA. A ribozyme having specificity for an LSO-encoding nucleic acid can be designed based upon the nucleotide sequence of an LSO cDNA disclosed herein (*i.e.*, SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an LSO-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, LSO mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, LSO gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the LSO (*e.g.*, the LSO promoter and/or enhancers; *e.g.*, nucleotides 1-94 of SEQ ID NO:1) to form triple helical structures that prevent transcription of the LSO gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6): 569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the LSO nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al. Proc. Natl. Acad. Sci.* 93: 14670-675.

PNAs of LSO nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of LSO nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of LSO can be modified, (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of LSO nucleic acid molecules can be generated which may combine the advantageous properties of PNA

and DNA. Such chimeras allow DNA recognition enzymes, (*e.g.*, RNAse H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

Alternatively, the expression characteristics of an endogenous LSO gene within a cell line or microorganism may be modified by inserting a heterologous DNA regulatory element into the genome of a stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous LSO gene. For example, an endogenous LSO gene which is normally "transcriptionally silent", *i.e.*, an LSO gene which is normally not expressed, or is expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell

line or microorganism. Alternatively, a transcriptionally silent, endogenous LSO gene may be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with an endogenous LSO gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described, *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

II. Isolated LSO Proteins and Anti-LSO Antibodies

One aspect of the invention pertains to isolated LSO proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-LSO antibodies. In one embodiment, native LSO proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, LSO proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an LSO protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the LSO protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of LSO protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of LSO protein having less than about 30% (by dry weight) of non-LSO protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-LSO protein, still more preferably less than about 10% of non-LSO protein, and most preferably less than about 5% non-LSO protein. When the LSO protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of LSO protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of LSO protein having less than about 30% (by dry weight) of chemical precursors or non-LSO chemicals, more preferably less than about 20% chemical precursors or non-LSO chemicals, still more preferably less than about 10% chemical precursors or non-LSO chemicals, and most preferably less than about 5% chemical precursors or non-LSO chemicals.

As used herein, a "biologically active portion" of an LSO protein includes a fragment of an LSO protein which participates in an interaction between an LSO molecule and a non-LSO molecule. Biologically active portions of an LSO protein include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the LSO protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length LSO proteins, and exhibit at least one activity of an LSO protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the LSO protein, *e.g.*, modulating membrane excitability. A biologically active portion of an LSO protein can be a polypeptide which is, for example, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 725, 750 or more amino acids in length. Biologically active portions of an LSO protein can be used as targets for developing agents which modulate an LSO mediated activity, *e.g.*, a proliferative response.

It is to be understood that a preferred biologically active portion of an LSO protein of the present invention may contain at least one or more of the following motifs or domains: a signal peptide, an N-glycosylation site, a lysyl oxidase domain, an LSO signature motif, and/or a scavenger receptor cysteine-rich domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native LSO protein.

In a preferred embodiment, the LSO protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the LSO protein is substantially identical to SEQ ID NO:2, and retains the functional activity of the protein of SEQ ID NO:2, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as

described in detail in subsection I above. Accordingly, in another embodiment, the LSO protein is a protein which comprises an amino acid sequence at least about 42%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the LSO amino acid sequence of SEQ ID NO:2 having 328 amino acid residues, at least 50, preferably at least 100, more preferably at least 200, even more preferably at least 300, at least 400, at least 500, at least 600, and even more preferably at least 700 or more amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at

http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*Comput. Appl. Biosci.*, 4: 11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to LSO nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 100, wordlength = 3 to obtain amino acid sequences homologous to LSO protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17): 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also provides LSO chimeric or fusion proteins. As used herein, an LSO "chimeric protein" or "fusion protein" comprises an LSO polypeptide operatively linked to a non-LSO polypeptide. An "LSO polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an LSO molecule, whereas a "non-LSO polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the LSO protein, *e.g.*, a protein which is different from the LSO protein and which is derived from the same or a different organism. Within an LSO fusion protein the LSO polypeptide can correspond to all or a portion of an LSO protein. In a preferred embodiment, an LSO fusion protein comprises at least one biologically active portion of an LSO protein. In another preferred embodiment, an LSO fusion protein comprises at least two biologically active portions of an LSO protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the LSO polypeptide and the non-LSO polypeptide are fused in-frame to

each other. The non-LSO polypeptide can be fused to the N-terminus or C-terminus of the LSO polypeptide.

For example, in one embodiment, the fusion protein is a GST-LSO fusion protein in which the LSO sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant LSO.

In another embodiment, the fusion protein is an LSO protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of LSO can be increased through use of a heterologous signal sequence.

The LSO fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The LSO fusion proteins can be used to affect the bioavailability of an LSO substrate. Use of LSO fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding an LSO protein; (ii) mis-regulation of the LSO gene; and (iii) aberrant post-translational modification of an LSO protein.

Moreover, the LSO-fusion proteins of the invention can be used as immunogens to produce anti-LSO antibodies in a subject, to purify LSO ligands and in screening assays to identify molecules which inhibit the interaction of LSO with an LSO substrate.

Preferably, an LSO chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An LSO-

encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the LSO protein.

The present invention also pertains to variants of the LSO proteins which function as either LSO agonists (mimetics) or as LSO antagonists. Variants of the LSO proteins can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of an LSO protein. An agonist of the LSO proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of an LSO protein. An antagonist of an LSO protein can inhibit one or more of the activities of the naturally occurring form of the LSO protein by, for example, competitively modulating an LSO-mediated activity of an LSO protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the LSO protein.

In one embodiment, variants of an LSO protein which function as either LSO agonists (mimetics) or as LSO antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of an LSO protein for LSO protein agonist or antagonist activity. In one embodiment, a variegated library of LSO variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of LSO variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential LSO sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of LSO sequences therein. There are a variety of methods which can be used to produce libraries of potential LSO variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential LSO sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of an LSO protein coding sequence can be used to generate a variegated population of LSO fragments for screening and subsequent selection of variants of an LSO protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an LSO coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the LSO protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of LSO proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify LSO variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3): 327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated LSO library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, a neuronal cell line, which ordinarily responds to an LSO ligand in a particular LSO ligand-dependent manner. The transfected cells are then contacted with an LSO ligand and the effect of expression of the mutant on, *e.g.*, membrane excitability of LSO can be detected. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the LSO ligand, and the individual clones further characterized.

An isolated LSO protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind LSO using standard techniques for polyclonal and monoclonal antibody preparation. A full-length LSO protein can be used or, alternatively, the invention provides antigenic peptide fragments of LSO for use as immunogens. The antigenic peptide of LSO comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of LSO such that an antibody raised against the peptide forms a specific immune complex with the LSO protein. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of LSO that are located on the surface of the protein, *e.g.*, hydrophilic regions, as well as regions with high antigenicity (see Figure 2).

An LSO immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed LSO protein or a chemically synthesized LSO polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic LSO preparation induces a polyclonal anti-LSO antibody response.

Accordingly, another aspect of the invention pertains to anti-LSO antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as an LSO. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind LSO molecules. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of LSO. A monoclonal antibody composition thus typically displays a single binding affinity for a particular LSO protein with which it immunoreacts.

Polyclonal anti-LSO antibodies can be prepared as described above by immunizing a suitable subject with an LSO immunogen. The anti-LSO antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized LSO. If desired, the antibody molecules directed against LSO can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-LSO antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem* .255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an LSO immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds LSO.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-LSO monoclonal antibody (see, *e.g.*, G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al. Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of

the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind LSO, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-LSO antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with LSO to thereby isolate immunoglobulin library members that bind LSO. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrad *et al.* (1991) *Bio/Technology* 9:1373-1377;

Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al. Nature* (1990) 348:552-554.

Additionally, recombinant anti-LSO antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-LSO antibody (*e.g.*, monoclonal antibody) can be used to isolate LSO by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-LSO antibody can facilitate the purification of natural LSO from cells and of recombinantly produced LSO expressed in host cells. Moreover, an anti-LSO antibody can be used to detect LSO protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the LSO protein. Anti-LSO antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include

streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an LSO protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of

interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, LSO proteins, mutant forms of LSO proteins, fusion proteins, and the like).

The recombinant expression vectors of the invention can be designed for expression of LSO proteins in prokaryotic or eukaryotic cells. For example, LSO proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein

from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in LSO activity assays, (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies specific for LSO proteins, for example. In a preferred embodiment, an LSO fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the LSO expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, LSO proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-

regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to LSO mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which an LSO nucleic acid molecule of the invention is introduced, *e.g.*, an LSO nucleic acid molecule within a recombinant expression vector or an LSO nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, an LSO protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an LSO protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an LSO protein. Accordingly, the invention further provides methods for producing an LSO protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding an LSO protein has been introduced) in a suitable medium such that an LSO protein is produced. In another embodiment, the method further comprises isolating an LSO protein from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which LSO-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous LSO sequences have been introduced into their genome or

homologous recombinant animals in which endogenous LSO sequences have been altered. Such animals are useful for studying the function and/or activity of an LSO and for identifying and/or evaluating modulators of LSO activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous LSO gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing an LSO-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The LSO cDNA sequence of SEQ ID NO:1 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human LSO gene, such as a mouse or rat LSO gene, can be used as a transgene. Alternatively, an LSO gene homologue, such as another LSO family member, can be isolated based on hybridization to the LSO cDNA sequences of SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____ (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to an LSO transgene to direct expression of an LSO protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory

Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of an LSO transgene in its genome and/or expression of LSO mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding an LSO protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an LSO gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the LSO gene. The LSO gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:3), but more preferably, is a non-human homologue of a human LSO gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1). For example, a mouse LSO gene can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous LSO gene in the mouse genome. In a preferred embodiment, the homologous recombination nucleic acid molecule is designed such that, upon homologous recombination, the endogenous LSO gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous LSO gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous LSO protein). In the homologous recombination nucleic acid molecule, the altered portion of the LSO gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the LSO gene to allow for homologous recombination to occur between the exogenous LSO gene carried by the homologous recombination nucleic acid molecule and an endogenous LSO gene in a cell, *e.g.*, an embryonic stem cell. The additional flanking LSO nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the homologous recombination nucleic acid molecule (see, *e.g.*, Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, *e.g.*, an embryonic stem cell line (*e.g.*, by

electroporation) and cells in which the introduced LSO gene has homologously recombined with the endogenous LSO gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, *e.g.*, vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, *e.g.*, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte

is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The LSO nucleic acid molecules, fragments of LSO proteins, and anti-LSO antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a fragment of an LSO protein or an anti-LSO antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and

used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected

cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e.*,. including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be

administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and

doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response; the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon *et al.*, "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin *et al.* (eds.), pp. 303-16 (Academic Press 1985), and Thorpe *et al.*, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982). Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which

the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, an LSO protein of the invention has one or more of the following activities: 1) it modulates cellular growth, proliferation, or differentiation, 2) it modulates skin formation or elasticity 3) it modulates bone formation or structure; 4) it modulates muscle formation or elasticity; 5) it modulates cartilage formation or structure; and 6) it modulates tumor formation and, thus, may be used to 1) modulate cellular growth, proliferation, or differentiation, 2) modulate skin formation or elasticity 3) modulate bone formation or structure; 4) modulate muscle formation or elasticity; 5) modulate cartilage formation or structure; and 6) modulate tumor formation

The isolated nucleic acid molecules of the invention can be used, for example, to express LSO protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect LSO mRNA (*e.g.*, in a biological sample) or a genetic alteration in an LSO gene, and to modulate LSO activity, as described further below. The LSO proteins can be used to treat disorders characterized by insufficient or excessive production of an LSO substrate or production of LSO inhibitors. In addition, the LSO proteins can be used to screen for naturally occurring LSO substrates, to screen for drugs or compounds which modulate LSO activity, as well as to treat disorders characterized by insufficient or excessive production of LSO protein or production of LSO protein forms which have decreased, aberrant or unwanted activity compared to LSO wild type protein (*e.g.*, lysyl oxidase-associated disorders).

In a preferred embodiment, the LSO molecules of the invention are useful for catalyzing the covalent cross-linking of the molecular units of elastin or collagen. As such, these molecules may be employed in small or large-scale synthesis of either elastin or collagen, or in chemical processes that require the production of these compounds. Such processes are known in the art (see, *e.g.*, Ullmann *et al.* (1999) Ullmann's Encyclopedia of Industrial Chemistry, 6th ed. VCH: Weinheim; Gutcho (1983) Chemicals by Fermentation. Park ridge, NJ: Noyes Data Corporation (ISBN 0818805086); Rehm *et al.* (eds.) (1993) Biotechnology, 2nd ed. VCH: Weinheim; and Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology. New York: John Wiley & Sons, and references contained therein.)

The isolated nucleic acid molecules of the invention can be used, for example, to express LSO protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect LSO mRNA (*e.g.*, in a biological sample) or a genetic alteration in an LSO gene, and to modulate LSO activity, as described further below. The LSO proteins can be used to treat disorders characterized by insufficient or excessive production of an LSO substrate or production of LSO inhibitors. In addition, the LSO proteins can be used to screen for naturally occurring LSO substrates, to screen for drugs or compounds which modulate LSO activity, as well as to treat disorders characterized by insufficient or excessive production of LSO protein or production of LSO protein forms which have decreased, aberrant or unwanted activity compared to LSO wild type protein (*e.g.*, lysyl oxidase-associated disorders, such as muscular disorders (*e.g.*, paralysis, muscle weakness (*e.g.*, ataxia, myotonia, and myokymia), muscular dystrophy (*e.g.*, Duchenne muscular dystrophy or myotonic dystrophy), spinal muscular atrophy, congenital myopathies, central core disease, rod myopathy, central nuclear myopathy, Lambert-Eaton syndrome, denervation, and infantile spinal muscular atrophy (Werdnig-Hoffman disease); cellular growth, differentiation, or migration disorders (*e.g.*, cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; neuronal deficiencies resulting from impaired neural induction and patterning); disorders of bone formation and resorption (*e.g.*, osteoporosis, osteochondrosis, and osteopetrosis); disorders of skin formation and elasticity (*e.g.*, cutis laxa and Ehlers-Danlos type V syndrome); or disorders of cartilage formation and structure (*e.g.*, chondromalacia and polychondritis). Moreover, the anti-LSO antibodies

of the invention can be used to detect and isolate LSO proteins, regulate the bioavailability of LSO proteins, and modulate LSO activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which interact with or bind to LSO proteins, have a stimulatory or inhibitory effect on, for example, LSO expression or LSO activity, or have a stimulatory or inhibitory effect on, for example, the availability of LSO substrate.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of an LSO protein or polypeptide or biologically active portion thereof (*e.g.*, subunits of elastin or collagen, or compounds which are structurally related thereto). In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of an LSO protein or polypeptide or biologically active portion thereof (*e.g.*, copper ions, lysyltyrosine quinone, trihydroxyphenylalanaine quinone, or other cofactors, or inhibitory molecules). The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses an LSO protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate LSO activity is determined. Determining the ability of the test compound to modulate LSO activity can be accomplished by monitoring, for example, the production of one or more specific LSO substrates or products in a cell which expresses LSO (see, *e.g.*, Saada *et al.* (2000) *Biochem Biophys. Res.Commun.* 269: 382-386). The cell, for example, can be of mammalian origin. The ability of the test compound to modulate LSO binding to a substrate (*e.g.*, subunits of elastin or collagen) or to bind to LSO can also be determined. Determining the ability of the test compound to modulate LSO binding to a substrate can be accomplished, for example, by coupling the LSO substrate or test compound with a radioisotope or fluorigenic label such that binding of the LSO substrate or test compound to LSO can be determined by detecting the labeled LSO substrate in a complex. Alternatively, LSO or a test compound could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate LSO binding to an LSO substrate in a complex. Determining the ability of the test compound to bind LSO can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic or otherwise detectable label such that binding of the compound to LSO can be determined by detecting the labeled LSO compound in a complex. For example, compounds (*e.g.*, LSO substrates) can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound (*e.g.*, an LSO substrate) to interact with LSO without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with LSO without the labeling of either the compound or the LSO. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and LSO.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing an LSO target molecule (*e.g.*, an LSO substrate) with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the LSO target molecule. Determining the ability of the test compound to modulate the activity of an LSO target molecule can be accomplished, for example, by determining the ability of the LSO protein to bind to or interact with the LSO target molecule.

Determining the ability of the LSO protein, or a biologically active fragment thereof, to bind to or interact with an LSO target molecule (*e.g.*, a substrate or inhibitor) can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the LSO protein to bind to or interact with an LSO target molecule can be accomplished by determining the activity or availability of the target molecule. For example, a target-regulated cellular activity, such as a biosynthetic pathway which requires the participation of the target molecule (e.g., elastin or collagen synthesis), may be monitored.

In yet another embodiment, an assay of the present invention is a cell-free assay in which an LSO protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to associate with, to bind to, or to serve as a substrate for the LSO protein or biologically active portion thereof is determined. Preferred biologically active portions of the LSO proteins to be used in assays of the present invention include fragments which participate in interactions with non-LSO molecules, *e.g.*, fragments with high surface probability scores (see Figure 2). Binding of the test compound to the LSO protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the LSO

protein or biologically active portion thereof with a known compound which interacts with LSO to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an LSO protein, wherein determining the ability of the test compound to interact with an LSO protein comprises determining the ability of the test compound to preferentially bind to or interact with LSO or a biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which an LSO protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the LSO protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of an LSO protein can be accomplished, for example, by determining the ability of the LSO protein to bind to or associate with an LSO target molecule by one of the methods described above for determining direct binding. Determining the ability of the LSO protein to bind to an LSO target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of an LSO protein can be accomplished by determining the ability of the LSO protein to further modulate the activity of a downstream effector of an LSO target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting an LSO protein or biologically active portion thereof with a known compound which binds the LSO protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the LSO protein, wherein determining the ability of the test compound to interact with the LSO protein comprises determining the ability of the LSO protein to preferentially bind to or

catalyze the oxidation of peptidyl lysine in the target substrate (*e.g.*, subunits of elastin or collagen).

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either LSO or its target molecule to facilitate separation of complexed from uncomplexed forms of either of the interactants, as well as to accommodate automation of the assay. Binding of a test compound to an LSO protein, or interaction of an LSO protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows the LSO protein to be bound to a matrix. For example, glutathione-S-transferase/LSO fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or LSO protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of LSO binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, an LSO protein can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated LSO protein can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with LSO protein but which do not interfere with binding of the LSO protein to its target molecule can be derivatized to the wells of the plate, and unbound target or LSO protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the LSO protein or target

molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the LSO protein.

In another embodiment, modulators of LSO expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of LSO mRNA or protein in the cell is determined. The level of expression of LSO mRNA or protein in the presence of the candidate compound is compared to the level of expression of LSO mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of LSO expression based on this comparison. For example, when expression of LSO mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of LSO mRNA or protein expression. Alternatively, when expression of LSO mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of LSO mRNA or protein expression. The level of LSO mRNA or protein expression in the cells can be determined by methods described herein for detecting LSO mRNA or protein.

In yet another aspect of the invention, the LSO proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with LSO ("LSO-binding proteins" or "LSO-6-bp") and are involved in LSO activity. Such LSO-binding proteins are also likely to be involved in the propagation of signals by the LSO proteins or LSO targets as, for example, downstream elements of an LSO-mediated signaling pathway. Alternatively, such LSO-binding proteins are likely to be LSO inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for an LSO protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If

the "bait" and the "prey" proteins are able to interact, *in vivo,* forming an LSO-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with the LSO protein.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of an LSO protein can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for cellular transformation and/or tumorigenesis.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, an LSO modulating agent, an antisense LSO nucleic acid molecule, an LSO-specific antibody, an LSO substrate or an LSO-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the LSO nucleotide sequences, described herein, can be used to map the location of the LSO genes on a chromosome. The mapping of the LSO sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, LSO genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the LSO nucleotide sequences. Computer analysis of the LSO sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the LSO sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the LSO nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map an LSO sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-

screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature,* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the LSO gene can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for

structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The LSO sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the LSO nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The LSO nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The

noncoding sequences of SEQ ID NO:1 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as that in SEQ ID NO:3, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from LSO nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of LSO Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the LSO nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases.

The LSO nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, thymus or brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such LSO probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, LSO primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining LSO protein and/or nucleic acid expression as well as LSO activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted LSO expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with LSO protein, nucleic acid expression or activity. For example, mutations in an LSO gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby phophylactically treat an individual prior to the onset of a disorder characterized by or associated with LSO protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of LSO in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of LSO protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of

detecting LSO protein or nucleic acid (*e.g.*, mRNA, or genomic DNA) that encodes LSO protein such that the presence of LSO protein or nucleic acid is detected in the biological sample. A preferred agent for detecting LSO mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to LSO mRNA or genomic DNA. The nucleic acid probe can be, for example, the LSO nucleic acid set forth in SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to LSO mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting LSO protein is an antibody capable of binding to LSO protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')2) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect LSO mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of LSO mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of LSO protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of LSO genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of LSO protein include introducing into a subject a labeled anti-LSO antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting LSO protein, mRNA, or genomic DNA, such that the presence of LSO protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of LSO protein, mRNA or genomic DNA in the control sample with the presence of LSO protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of LSO in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting LSO protein or mRNA in a biological sample; means for determining the amount of LSO in the sample; and means for comparing the amount of LSO in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect LSO protein or nucleic acid.

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted LSO expression or activity. As used herein, the term "aberrant" includes an LSO expression or activity which deviates from the wild type LSO expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type developmental pattern of expression or the subcellular pattern of expression. For example, aberrant LSO expression or activity is intended to include the cases in which a mutation in the LSO gene causes the LSO gene to be under-expressed or over-expressed and situations in which such mutations result in a non-functional LSO protein or a protein which does not function in a wild-type fashion, *e.g.*, a protein which does not interact with an LSO substrate, or one which interacts with a non-LSO substrate. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a

biological response such as cellular proliferation. For example, the term unwanted includes an LSO expression or activity which is undesirable in a subject.

The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with a misregulation in LSO protein activity or nucleic acid expression, such as a cellular proliferation, growth, differentiation, or migration disorder, a muscular disorder, a disorder of bone formation or structure, a disorder of cartilage formation or structure, a disorder of skin elasticity or formation, or a tumor formation disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in LSO protein activity or nucleic acid expression, such as a cellular proliferation, growth, differentiation, or migration disorder, a muscular disorder, a disorder of bone formation or structure, a disorder of cartilage formation or structure, a disorder of skin elasticity or formation, or a tumor formation disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted LSO expression or activity in which a test sample is obtained from a subject and LSO protein or nucleic acid (*e.g.*, mRNA or genomic DNA) is detected, wherein the presence of LSO protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted LSO expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, cerebrospinal fluid or serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted LSO expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a cellular proliferation, growth, differentiation, or migration disorder, a muscular disorder, a disorder of bone formation or structure, a disorder of cartilage formation or structure, a disorder of skin elasticity or formation, or a tumor formation disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted LSO expression or activity in which a test sample

is obtained and LSO protein or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of LSO protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted LSO expression or activity).

The methods of the invention can also be used to detect genetic alterations in an LSO gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in LSO protein activity or nucleic acid expression, such as a cellular proliferation, growth, differentiation, or migration disorder, a muscular disorder, a disorder of bone formation or structure, a disorder of cartilage formation or structure, a disorder of skin elasticity or formation, or a tumor formation disorder. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding an LSO-protein, or the mis-expression of the LSO gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from an LSO gene; 2) an addition of one or more nucleotides to an LSO gene; 3) a substitution of one or more nucleotides of an LSO gene, 4) a chromosomal rearrangement of an LSO gene; 5) an alteration in the level of a messenger RNA transcript of an LSO gene, 6) aberrant modification of an LSO gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of an LSO gene, 8) a non-wild type level of an LSO-protein, 9) allelic loss of an LSO gene, and 10) inappropriate post-translational modification of an LSO-protein. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in an LSO gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in an LSO gene (see Abravaya *et al.* (1995) *Nucleic Acids Res* .23:675-682). This method can include the steps of

collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to an LSO gene under conditions such that hybridization and amplification of the LSO gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an LSO gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in LSO can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in LSO can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by

making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the LSO gene and detect mutations by comparing the sequence of the sample LSO with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the LSO gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type LSO sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.*

(1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in LSO cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on an LSO sequence, *e.g.*, a wild-type LSO sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in LSO genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control LSO nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of

high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an LSO gene.

Furthermore, any cell type or tissue in which LSO is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of an LSO protein (*e.g.*, the modulation of cell proliferation and/or migration) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase LSO gene expression, protein levels, or upregulate LSO activity, can be monitored in clinical trials of subjects exhibiting decreased LSO gene expression, protein levels, or downregulated LSO activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease LSO gene expression, protein levels, or downregulate LSO activity, can be monitored in clinical trials of subjects exhibiting increased LSO gene expression, protein levels, or upregulated LSO activity. In such clinical trials, the expression or activity of an LSO gene, and preferably, other genes that have been implicated in, for example, an LSO-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including LSO, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates LSO activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on LSO-associated disorders (*e.g.*, disorders characterized by deregulated cell proliferation and/or migration), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of LSO and other genes implicated in the LSO-associated disorder, respectively. The levels of gene expression (*e.g.*, a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of LSO or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) including the steps of (i)

obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of an LSO protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the LSO protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the LSO protein, mRNA, or genomic DNA in the pre-administration sample with the LSO protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of LSO to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of LSO to lower levels than detected, *i.e.* to decrease the effectiveness of the agent. According to such an embodiment, LSO expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

D. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted LSO expression or activity, *e.g.*, a lysyl oxidase-associated disorder such as a cellular proliferation, growth, differentiation, or migration disorder, a muscular disorder, a disorder of bone formation or structure, a disorder of cartilage formation or structure, a disorder of skin elasticity or formation, or a tumor formation disorder. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics.

As used herein, "treatment" of a subject includes the application or administration of a therapeutic agent to a subject, or application or administration of a therapeutic agent to a cell or tissue from a subject, who has a diseases or disorder, has a symptom of a disease or disorder, or is at risk of (or susceptible to) a disease or disorder, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve, or affect the disease or disorder, the symptom of the disease or disorder, or the risk of (or susceptibility to) the disease or disorder. As used herein, a "therapeutic agent" includes,

but is not limited to, small molecules, peptides, polypeptides, antibodies, ribozymes, and antisense oligonucleotides.

"Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the LSO molecules of the present invention or LSO modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted LSO expression or activity, by administering to the subject an LSO or an agent which modulates LSO expression or at least one LSO activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted LSO expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the LSO aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of LSO aberrancy, for example, an LSO, LSO agonist or LSO antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating LSO expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with an LSO or agent that modulates one or more of the activities of LSO protein activity

associated with the cell. An agent that modulates LSO protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of an LSO protein (*e.g.*, an LSO substrate), an LSO antibody, an LSO agonist or antagonist, a peptidomimetic of an LSO agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more LSO activities. Examples of such stimulatory agents include active LSO protein and a nucleic acid molecule encoding LSO that has been introduced into the cell. In another embodiment, the agent inhibits one or more LSO activities. Examples of such inhibitory agents include antisense LSO nucleic acid molecules, anti-LSO antibodies, and LSO inhibitors. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of an LSO protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) LSO expression or activity. In another embodiment, the method involves administering an LSO protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted LSO expression or activity.

Stimulation of LSO activity is desirable in situations in which LSO is abnormally downregulated and/or in which increased LSO activity is likely to have a beneficial effect. Likewise, inhibition of LSO activity is desirable in situations in which LSO is abnormally upregulated and/or in which decreased LSO activity is likely to have a beneficial effect.

3. Pharmacogenomics

The LSO molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on LSO activity (*e.g.*, LSO gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) LSO-associated disorders (*e.g.*, proliferative disorders, muscular disorders, bone disorders, skin disorders, cartilage disorders, or tumor disorders) associated with aberrant or unwanted LSO activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship

between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer an LSO molecule or LSO modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with an LSO molecule or LSO modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp.Pharmacol. Physiol.* 23(10-11): 983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may

be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug target is known (*e.g.*, an LSO protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, an LSO molecule or LSO modulator of the present

invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an LSO molecule or LSO modulator, such as a modulator identified by one of the exemplary screening assays described herein.

E. Electronic Apparatus Readable Media and Arrays

Electronic apparatus readable media comprising LSO sequence information is also provided. As used herein, "LSO sequence information" refers to any nucleotide and/or amino acid sequence information particular to the LSO molecules of the present invention, including but not limited to full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences, polymorphic sequences including single nucleotide polymorphisms (SNPs), epitope sequences, and the like. Moreover, information "related to" said LSO sequence information includes detection of the presence or absence of a sequence (*e.g.*, detection of expression of a sequence, fragment, polymorphism, etc.), determination of the level of a sequence (*e.g.*, detection of a level of expression, for example, a quantitative detection), detection of a reactivity to a sequence (*e.g.*, detection of protein expression and/or levels, for example, using a sequence-specific antibody), and the like. As used herein, "electronic apparatus readable media" refers to any suitable medium for storing, holding, or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact discs; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon LSO sequence information of the present invention.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatuses; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the LSO sequence information.
A variety of software programs and formats can be used to store the sequence information on the electronic apparatus readable medium. For example, the sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, represented in the form of an ASCII file, or stored in a database application, such as DB2, Sybase, Oracle, or the like, as well as in other forms. Any number of dataprocessor structuring formats (*e.g.*, text file or database) may be employed in order to obtain or create a medium having recorded thereon the LSO sequence information.

By providing LSO sequence information in readable form, one can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the sequence information in readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a lysyl oxidase associated disorder or a pre-disposition to lysyl oxidase-associated disorder, wherein the method comprises the steps of determining LSO sequence information associated with the subject and based on the LSO sequence information, determining whether the subject has a lysyl oxidase-associated disorder or a pre-disposition to a lysyl oxidase-associated disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a lysyl oxidase-associated disorder or a pre-disposition to a lysyl oxidase-associated disorder wherein the method comprises the steps of determining LSO sequence information associated with the subject, and based on the LSO sequence information, determining whether the subject has a lysyl oxidase-associated disorder or a pre-disposition to a lysyl oxidase-associated disorder, and/or recommending a particular treatment for the disease, disorder or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

The present invention also provides in a network, a method for determining whether a subject has a lysyl oxidase-associated disorder or a pre-disposition to a lysyl oxidase-associated disorder associated with LSO, said method comprising the steps of receiving LSO sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to LSO and/or a lysyl oxidase-associated disorder, and based on one or more of the phenotypic information, the LSO information (*e.g.*, sequence information and/or information related thereto), and the acquired information, determining whether the subject has a lysyl oxidase-associated disorder or a pre-disposition to a lysyl oxidase-associated disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The present invention also provides a business method for determining whether a subject has lysyl oxidase-associated disorder or a pre-disposition to a lysyl oxidase-associated disorder, said method comprising the steps of receiving information related to LSO (*e.g.*, sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to LSO and/or related to a lysyl oxidase-associated disorder, and based on one or more of the phenotypic information, the LSO information, and the acquired information, determining whether the subject has a lysyl oxidase-associated disorder or a pre-disposition to a lysyl oxidase-associated disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The invention also includes an array comprising a LSO sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be LSO. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a lysyl oxidase-associated disorder, progression of a lysyl oxidase-associated disorder, and processes associated with the lysyl oxidase-associated disorder.

The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, ascertaining the effect of LSO expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including LSO) that could serve as a molecular target for diagnosis or therapeutic intervention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures and Sequence Listing, are incorporated herein by reference.

EXAMPLES

EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMANLSO cDNA

In this example, the identification and characterization of the gene encoding human LSO (clone Fbh47765) is described.

Isolation of the LSO cDNA

The invention is based, at least in part, on the discovery of a human gene encoding a novel protein, referred to herein as human LSO. The entire sequence of human clone Fbh47765, was determined and found to contain an open reading frame termed human "LSO".

The nucleotide sequence encoding the human LSO protein is shown in Figure 2 and is set forth as SEQ ID NO:1. The protein encoded by this nucleic acid comprises about 756 amino acids and has the amino acid sequence shown in Figure 1 and set forth as SEQ ID NO: 2. The coding region (open reading frame) of SEQ ID NO:1 is set forth as SEQ ID NO:3. Clone Fbh55158, comprising the coding region of human LSO, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

Analysis of the Human LSO Molecules

The amino acid sequence of human LSO was analyzed using the program PSORT (http://www. psort.nibb.ac.jp) to predict the localization of the protein within the cell. This program assesses the presence of different targeting and localization amino acid sequences within the query sequence. The results of the analyses show the likelihood of the human LSO (SEQ ID NO:2) being localized, for example, extracellularly, in the mitochondrion, in vacuoles, and in the endoplasmic reticulum.

An analysis of the amino acid sequence of human LSO using the Signal P program (Henrik, *et al.* (1997) *Protein Engineering* 10:1-6), identified the presence of a signal peptide from amino acids 1-24.

A search was also performed against the Prosite database resulting in the identification of two N-glycosylation sites in the amino acid sequence of human LSO (SEQ ID NO:2) at about residues 198-201 and 629-632.

A search of the amino acid sequence of human LSO was also performed against the HMM database. This search resulted in the identification of a "lysyl oxidase domain" in the amino acid sequence of human LSO (SEQ ID NO:2) at about residues 533-736 (score = 513.1). This search further resulted in the identification of four "scavenger receptor cysteine-rich domains" in the amino acid sequence of human LSO (SEQ ID NO:2) at about residues 37-133 (score = 98.1), 169-287 (score = 30.4), 314-411 (score = 115.8), and 424-529 (score = 46.3).

A search of the amino acid sequence of human LSO was also performed against the ProDom database, resulting in the identification of a "lysyl oxidase protein-lysine precursor signal 6-oxidase oxireductase copper glycoprotein homolog" domain at amino acid residues 533-731 of human LSO.

Tissue Distribution of LSO mRNA

This example describes the tissue distribution of human LSO mRNA, as determined by Northern analysis, by Polymerase Chain Reaction (PCR) on cDNA libraries using oligonucleotide primers based on the human LSO sequence, or by *in situ* analysis.

Northern blot hybridizations with the various RNA samples are performed under standard conditions and washed under stringent conditions, *i.e.*, 0.2xSSC at 65°C. The DNA probe is radioactively labeled with ^{32}P-dCTP using the Prime-It kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing human

mRNA (MultiTissue Northern I and MultiTissue Northern II from Clontech, Palo Alto, CA) are probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Human LSO expression in normal tissues, *e.g.*, human tissues, is assessed by PCR using the Taqman ® system (PE Applied Biosystems) according to the manufacturer's instructions.

For *in situ* analysis, various tissues, *e.g.* tissues obtained from brain, are first frozen on dry ice. Ten-micrometer-thick sections of the tissues are postfixed with 4% formaldehyde in DEPC treated 1X phosphate- buffered saline at room temperature for 10 minutes before being rinsed twice in DEPC 1X phosphate-buffered saline and once in 0.1 M triethanolamine-HCl (pH 8.0). Following incubation in 0.25% acetic anhydride-0.1 M triethanolamine-HCl for 10 minutes, sections are rinsed in DEPC 2X SSC (1X SSC is 0.15M NaCl plus 0.015M sodium citrate). Tissue is then dehydrated through a series of ethanol washes, incubated in 100% chloroform for 5 minutes, and then rinsed in 100% ethanol for 1 minute and 95% ethanol for 1 minute and allowed to air dry.

Hybridizations are performed with ^{35}S-radiolabeled (5 X 10^7 cpm/ml) cRNA probes. Probes are incubated in the presence of a solution containing 600 mM NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.01% sheared salmon sperm DNA, 0.01% yeast tRNA, 0.05% yeast total RNA type X1, 1X Denhardt's solution, 50% formamide, 10% dextran sulfate, 100 mM dithiothreitol, 0.1% sodium dodecyl sulfate (SDS), and 0.1% sodium thiosulfate for 18 hours at 55°C.

After hybridization, slides are washed with 2X SSC. Sections are then sequentially incubated at 37°C in TNE (a solution containing 10 mM Tris-HCl (pH 7.6), 500 mM NaCl, and 1 mM EDTA), for 10 minutes, in TNE with 10μg of RNase A per ml for 30 minutes, and finally in TNE for 10 minutes. Slides are then rinsed with 2X SSC at room temperature, washed with 2X SSC at 50°C for 1 hour, washed with 0.2X SSC at 55°C for 1 hour, and 0.2X SSC at 60°C for 1 hour. Sections are then dehydrated rapidly through serial ethanol-0.3 M sodium acetate concentrations before being air dried and exposed to Kodak Biomax MR scientific imaging film for 24 hours and subsequently dipped in NB-2 photoemulsion and exposed at 4°C for 7 days before being developed and counter stained.

EXAMPLE 2: EXPRESSION OF RECOMBINANT HUMAN LSO PROTEIN IN BACTERIAL CELLS

In this example, human LSO is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, human LSO is fused to GST and this fusion polypeptide is expressed in *E. coli*, *e.g.*, strain PEB199. Expression of the GST-LSO fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

EXAMPLE 3: EXPRESSION OF RECOMBINANT HUMAN LSO PROTEIN IN COS CELLS

To express the human LSO gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire human LSO protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

To construct the plasmid, the human LSO DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the human LSO coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the human LSO coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the human LSO gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5α, SURE, available from Stratagene Cloning Systems, La

Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the human LSO-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the LSO polypeptide is detected by radiolabelling (^{35}S-methionine or ^{35}S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA-specific, a FLAG-specific, or a human LSO-specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA-specific, a FLAG-specific, or a human LSO-specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the human LSO coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the human LSO polypeptide is detected by radiolabelling and immunoprecipitation using a human LSO-specific monoclonal antibody.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1; and

(b) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:3.

2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2.

3. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____.

4. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2.

5. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1 or 3, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 50 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 60% identical to the amino acid sequence of SEQ ID NO:2; and

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment

comprises at least 15 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2.

6. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.

7. An isolated nucleic acid molecule comprising a nucleotide sequence which is complementary to the nucleotide sequence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.

9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

10. The vector of claim 9, which is an expression vector.

11. A host cell transfected with the expression vector of claim 10.

12. A method of producing a polypeptide comprising culturing the host cell of claim 11 in an appropriate culture medium to, thereby, produce the polypeptide.

13. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of SEQ ID NO:1 or 3 under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3;

d) a polypeptide comprising an amino acid sequence which is at least 60% identical to the amino acid sequence of SEQ ID NO:2.

14. The isolated polypeptide of claim 13 comprising the amino acid sequence of SEQ ID NO:2.

15. The polypeptide of claim 13, further comprising heterologous amino acid sequences.

16. An antibody which selectively binds to a polypeptide of claim 13.

17. A method for detecting the presence of a polypeptide of claim 13 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 13 in the sample.

18. The method of claim 17, wherein the compound which binds to the polypeptide is an antibody.

19. A kit comprising a compound which selectively binds to a polypeptide of claim 13 and instructions for use.

20. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

21. The method of claim 20, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

22. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

23. A method for identifying a compound which binds to a polypeptide of claim 13 comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

24. The method of claim 23, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for LSO activity.

25. A method for modulating the activity of a polypeptide of claim 13 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

26. A method for identifying a compound which modulates the activity of a polypeptide of claim 13 comprising:

a) contacting a polypeptide of claim 13 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

5

FIG. 1A

> Fbh47765Fl - Import - vector trimmed

```
CACGCGTCCGCGCTGCTCCGCCGCGGCGCCCGCCCAGCCCCGGACTGTCCGCGCTCCATCTGGTATCTTGGCCTCAGCT

                    M   A   W   S   P   P   A   T   L   F   L   F   L   L   L   L    16
GTCCTTGAAGTCACC ATG GCG TGG TCC CCA CCA GCC ACC CTC TTT CTG TTC CTG CTG CTG CTA   48

 G   Q   P   P   P   S   R   P   Q   S   L   G   T   T   K   L   R   L   V   G    36
GGC CAG CCC CCT CCC AGC AGG CCA CAG TCA CTG GGC ACC ACT AAG CTC CGG CTG GTG GGC  108

 P   E   S   K   P   E   E   G   R   L   E   V   L   H   Q   G   Q   W   G   T    56
CCA GAG AGC AAG CCA GAG GAG GGC CGC CTG GAG GTG CTG CAC CAG GGC CAG TGG GGC ACC  168

 V   C   D   D   N   F   A   I   Q   E   A   T   V   A   C   R   Q   L   G   F    76
GTG TGT GAT GAC AAC TTT GCT ATC CAG GAG GCC ACA GTG GCT TGC CGC CAG CTG GGC TTC  228

 E   A   A   L   T   W   A   H   S   A   K   Y   G   Q   G   E   G   P   I   W    96
GAA GCT GCC TTG ACC TGG GCC CAC AGT GCC AAG TAC GGC CAA GGG GAG GGA CCC ATC TGG  288

 L   D   N   V   R   C   V   G   T   E   S   S   L   D   Q   C   G   S   N   G   116
CTG GAC AAT GTG CGC TGT GTG GGC ACA GAG AGC TCC TTG GAC CAG TGC GGG TCT AAT GGC  348

 W   G   V   S   D   C   S   H   S   E   D   V   G   V   I   C   H   P   R   R   136
TGG GGA GTC AGT GAC TGC AGT CAC TCA GAA GAC GTA GGG GTG ATA TGC CAC CCC CGG CGC  408

 H   R   G   Y   L   S   E   T   V   S   N   A   L   G   P   Q   G   R   R   L   156
CAT CGT GGC TAC CTT TCT GAA ACT GTC TCC AAT GCC CTT GGG CCC CAG GGC CGG CGG CTG  468

 E   E   V   R   L   K   P   I   L   A   S   A   K   Q   H   S   P   V   T   E   176
GAG GAG GTG CGG CTC AAG CCC ATC CTT GCC AGT GCC AAG CAG CAT AGC CCA GTG ACC GAG  528
```

FIG. 1B

```
 G   A   V   E   V   K   Y   E   G   H   W   R   Q   V   C   D   Q   G   W   T   196
GGA GCC GTG GAG GTG AAG TAT GAG GGC CAC TGG CGG CAG GTG TGT GAC CAG GGC TGG ACC   588

 M   N   N   S   R   V   V   C   G   M   L   G   F   P   S   E   V   P   V   D   216
ATG AAC AAC AGC AGG GTG GTG TGC GGG ATG CTG GGC TTC CCC AGC GAG GTG CCT GTC GAC   648

 S   H   Y   Y   R   K   V   W   D   L   K   M   R   D   P   K   S   R   L   K   236
AGC CAC TAC TAC AGG AAA GTC TGG GAT CTG AAG ATG AGG GAC CCT AAG TCT AGG CTG AAG   708

 S   L   T   N   K   N   S   F   W   I   H   Q   V   T   C   L   G   T   E   P   256
AGC CTG ACG AAT AAG AAC TCC TTC TGG ATC CAC CAG GTC ACC TGC CTG GGG ACA GAG CCC   768

 H   M   A   N   C   Q   V   Q   V   A   P   A   R   G   K   L   R   P   A   C   276
CAC ATG GCC AAC TGC CAG GTG CAG GTG GCT CCA GCC CGG GGC AAG CTG CGG CCA GCC TGC   828

 P   G   G   M   H   A   V   V   S   C   V   A   G   P   H   F   R   P   P   K   296
CCA GGT GGC ATG CAC GCT GTG GTC AGC TGT GTG GCA GGG CCT CAC TTC CGC CCA CCG AAG   888

 T   K   P   Q   R   K   G   S   W   A   E   E   P   R   V   R   L   R   S   G   316
ACA AAG CCA CAA CGC AAA GGG TCC TGG GCA GAG GAG CCG AGG GTG CGC CTG CGC TCC GGG   948

 A   Q   V   G   E   G   R   V   E   V   L   M   N   R   Q   W   G   T   V   C   336
GCC CAG GTG GGC GAG GGC CGG GTG GAA GTG CTC ATG AAC CGC CAG TGG GGC ACG GTC TGT  1008

 D   H   R   W   N   L   I   S   A   S   V   V   C   R   Q   L   G   F   G   S   356
GAC CAC AGG TGG AAC CTC ATC TCT GCC AGT GTC GTG TGT CGT CAG CTG GGC TTT GGC TCT  1068

 A   R   E   A   L   F   G   A   R   L   G   Q   G   L   G   P   I   H   L   S   376
GCT CGG GAG GCC CTC TTT GGG GCC CGG CTG GGC CAA GGG CTA GGG CCC ATC CAC CTG AGT  1128
```

FIG. 1C

```
 E   V   R   C   R   G   Y   E   R   T   L   S   D   C   P   A   L   E   G   S   396
GAG GTG CGC TGC AGG GGA TAT GAG CGG ACC CTC AGC GAC TGC CCT GCC CTG GAA GGG TCC 1188

 Q   N   G   C   Q   H   E   N   D   A   A   V   R   C   N   V   P   N   M   G   416
CAG AAT GGT TGC CAA CAT GAG AAT GAT GCT GCT GTC AGG TGC AAT GTC CCT AAC ATG GGC 1248

 F   Q   N   Q   V   R   L   A   G   G   R   I   P   E   E   G   L   L   E   V   436
TTT CAG AAT CAG GTG CGC TTG GCT GGT GGG CGT ATC CCT GAG GAG GGG CTA TTG GAG GTG 1308

 Q   V   E   V   N   G   V   P   R   W   G   S   V   C   S   E   N   W   G   L   456
CAG GTG GAG GTG AAC GGG GTC CCA CGC TGG GGG AGC GTG TGC AGT GAA AAC TGG GGG CTC 1368

 T   E   A   M   V   A   C   R   Q   L   G   L   G   F   A   I   H   A   Y   K   476
ACC GAA GCC ATG GTG GCC TGC CGA CAG CTC GGC CTG GGT TTT GCC ATC CAT GCC TAC AAG 1428

 E   T   W   F   W   S   G   T   P   R   A   Q   E   V   V   M   S   G   V   R   496
GAA ACC TGG TTC TGG TCG GGG ACG CCA AGG GCC CAG GAG GTG GTG ATG AGT GGG GTG CGC 1488

 C   S   G   T   E   L   A   L   Q   Q   C   Q   R   H   G   P   V   H   C   S   516
TGC TCA GGC ACA GAG CTG GCC CTG CAG CAG TGC CAG AGG CAC GGG CCG GTG CAC TGC TCC 1548

 H   G   G   G   R   F   L   A   G   V   S   C   M   D   S   A   P   D   L   V   536
CAC GGT GGC GGG CGC TTC CTG GCT GGA GTC TCC TGC ATG GAC AGT GCA CCA GAC CTG GTG 1608

 M   N   A   Q   L   V   Q   E   T   A   Y   L   E   D   R   P   L   S   Q   L   556
ATG AAC GCC CAG CTA GTG CAG GAG ACG GCC TAC TTG GAG GAC CGC CCG CTC AGC CAG CTG 1668

 Y   C   A   H   E   E   N   C   L   S   K   S   A   D   H   M   D   W   P   Y   576
TAT TGT GCC CAC GAG GAG AAC TGC CTC TCC AAG TCT GCA GAT CAC ATG GAC TGG CCC TAC 1728
```

FIG. 1D

```
 G   Y   R   R   L   L   R   F   S   T   Q   I   Y   N   L   G   R   T   D   F   596
GGA TAC CGC CGC CTA TTG CGC TTC TCC ACA CAG ATC TAC AAT CTG GGC CGG ACT GAC TTT 1788

 R   P   K   T   G   R   D   S   W   V   W   H   Q   C   H   R   H   Y   H   S   616
CGT CCA AAG ACT GGA CGC GAT AGC TGG GTT TGG CAC CAG TGC CAC AGG CAT TAC CAC AGC 1848

 I   E   V   F   T   H   Y   D   L   L   T   L   N   G   S   K   V   A   E   G   636
ATT GAG GTC TTC ACC CAC TAC GAC CTC CTC ACT CTC AAT GGC TCC AAG GTG GCT GAG GGG 1908

 H   K   A   S   F   C   L   E   D   T   N   C   P   T   G   L   Q   R   R   Y   656
CAC AAG GCC AGC TTC TGT CTG GAG GAC ACA AAC TGC CCC ACA GGA CTG CAG CGG CGC TAC 1968

 A   C   A   N   F   G   E   Q   G   V   T   V   G   C   W   D   T   Y   R   H   676
GCA TGT GCC AAC TTT GGA GAA CAG GGA GTG ACT GTA GGC TGC TGG GAC ACC TAC CGG CAT 2028

 D   I   D   C   Q   W   V   D   I   T   D   V   G   P   G   N   Y   I   F   Q   696
GAC ATT GAT TGC CAG TGG GTG GAT ATC ACA GAT GTG GGC CCC GGG AAT TAT ATC TTC CAG 2088

 V   I   V   N   P   H   Y   E   V   A   E   S   D   F   S   N   N   M   L   Q   716
GTG ATT GTG AAC CCC CAC TAT GAA GTG GCA GAG TCA GAT TTC TCC AAC AAT ATG CTG CAG 2148

 C   R   C   K   Y   D   G   H   R   V   W   L   H   N   C   H   T   G   N   S   736
TGC CGC TGC AAG TAT GAT GGG CAC CGG GTC TGG CTG CAC AAC TGC CAC ACA GGG AAT TCA 2208

 Y   Y   A   N   A   E   L   S   L   E   Q   E   Q   R   L   R   N   N   L   I   756
TAC CCA GCC AAT GCA GAA CTC TCC CTG GAG CAG GAA CAG CGT CTC AGG AAC AAC CTC ATC 2268

 *                                                                               757
TGA                                                                             2271
```

FIG. 1E

AGCTGTCACTGCACACTCCTAGCTGCCGATACACCAGATACCTCAGCTTATTGGAGCCATGCCCTTCACAGAGTCC

CAACTCAGAGGAAAAGGGCCAGTGCCAAGGGGCACCAAGAACCTGCTCAGGAAGCCTTTTGATGGCAAGATCACCAATC

CAGATGGTATTGCTCCCTCAGGATGGCTCTGGGCCTGCCCCTAAGGGCCTGTGGCCTATGGAATATGTCCTCCAGGCTT

TGCTTAGCTGAGCTCCTCTTCTGTAAGGAAACCCAGTCATCCCTGAATCTTGCCACAGAGATCCGGGATTCAGGAGCTC

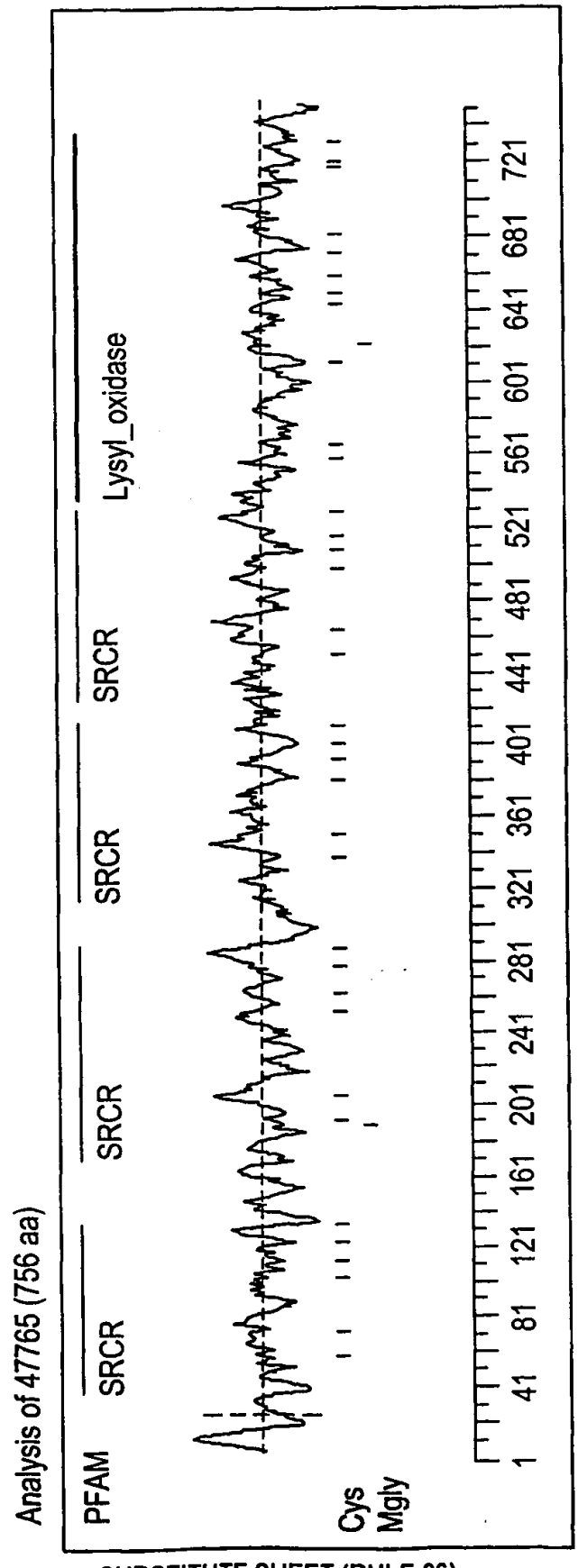
TCAGTTTCTTAGGGATGGACTATGGCCCAGTCCCCCATCTAAGTGGTGCTTTGCAAATGTCTTGGAGGAGTATAGGACA

GAGGACCAAAATACACAGCAGGTAGTGKTAGCTCTCTGCTAGGAGCTCAAAGCAACACAACTTGTATCAAAATCACAAC

TGGCAGAAAAACTGGGTGGATCCAATCCTTTCTTTCATCTGTGTATTTAAGAAYYAACCYTTAMACTCTGGTCTTTAGG

GGCCTTAMCTTTATTTWMCACAMAMAATNGGNGGTTTTATTATCYTTGGAAGCACAAA

FIG. 2



SEQUENCE LISTING

<110> MILLENNIUM PHARMACEUTICALS, INC. et al.

<120> 47765, A Novel Human Lysyl Oxidase and
Uses Thereof

<130> MNI-160PC

<150> 60/207,650
<151> 2000-05-26

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 2976
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (95)...(2362)

<221> misc_feature
<222> (1)...(2976)
<223> n = A,T,C or G

<400> 1
cacgcgtccg cgctgctccg ccgcggcgcc cgcccagccc cggactgtcc gcgctccatc 60
tggtatcttg gcctcagctg tccttgaagt cacc atg gcg tgg tcc cca cca gcc 115
Met Ala Trp Ser Pro Pro Ala
1 5

acc ctc ttt ctg ttc ctg ctg ctg cta ggc cag ccc cct ccc agc agg 163
Thr Leu Phe Leu Phe Leu Leu Leu Leu Gly Gln Pro Pro Pro Ser Arg
10 15 20

cca cag tca ctg ggc acc act aag ctc cgg ctg gtg ggc cca gag agc 211
Pro Gln Ser Leu Gly Thr Thr Lys Leu Arg Leu Val Gly Pro Glu Ser
25 30 35

aag cca gag gag ggc cgc ctg gag gtg ctg cac cag ggc cag tgg ggc 259
Lys Pro Glu Glu Gly Arg Leu Glu Val Leu His Gln Gly Gln Trp Gly
40 45 50 55

acc gtg tgt gat gac aac ttt gct atc cag gag gcc aca gtg gct tgc 307
Thr Val Cys Asp Asp Asn Phe Ala Ile Gln Glu Ala Thr Val Ala Cys
60 65 70

cgc cag ctg ggc ttc gaa gct gcc ttg acc tgg gcc cac agt gcc aag 355
Arg Gln Leu Gly Phe Glu Ala Ala Leu Thr Trp Ala His Ser Ala Lys
75 80 85

tac ggc caa ggg gag gga ccc atc tgg ctg gac aat gtg cgc tgt gtg 403
Tyr Gly Gln Gly Glu Gly Pro Ile Trp Leu Asp Asn Val Arg Cys Val
90 95 100

ggc aca gag agc tcc ttg gac cag tgc ggg tct aat ggc tgg gga gtc 451

```
Gly Thr Glu Ser Ser Leu Asp Gln Cys Gly Ser Asn Gly Trp Gly Val
    105                 110                 115

agt gac tgc agt cac tca gaa gac gta ggg gtg ata tgc cac ccc cgg   499
Ser Asp Cys Ser His Ser Glu Asp Val Gly Val Ile Cys His Pro Arg
120                 125                 130                 135

cgc cat cgt ggc tac ctt tct gaa act gtc tcc aat gcc ctt ggg ccc   547
Arg His Arg Gly Tyr Leu Ser Glu Thr Val Ser Asn Ala Leu Gly Pro
                140                 145                 150

cag ggc cgg cgg ctg gag gag gtg cgg ctc aag ccc atc ctt gcc agt   595
Gln Gly Arg Arg Leu Glu Glu Val Arg Leu Lys Pro Ile Leu Ala Ser
            155                 160                 165

gcc aag cag cat agc cca gtg acc gag gga gcc gtg gag gtg aag tat   643
Ala Lys Gln His Ser Pro Val Thr Glu Gly Ala Val Glu Val Lys Tyr
        170                 175                 180

gag ggc cac tgg cgg cag gtg tgt gac cag ggc tgg acc atg aac aac   691
Glu Gly His Trp Arg Gln Val Cys Asp Gln Gly Trp Thr Met Asn Asn
    185                 190                 195

agc agg gtg gtg tgc ggg atg ctg ggc ttc ccc agc gag gtg cct gtc   739
Ser Arg Val Val Cys Gly Met Leu Gly Phe Pro Ser Glu Val Pro Val
200                 205                 210                 215

gac agc cac tac tac agg aaa gtc tgg gat ctg aag atg agg gac cct   787
Asp Ser His Tyr Tyr Arg Lys Val Trp Asp Leu Lys Met Arg Asp Pro
                220                 225                 230

aag tct agg ctg aag agc ctg acg aat aag aac tcc ttc tgg atc cac   835
Lys Ser Arg Leu Lys Ser Leu Thr Asn Lys Asn Ser Phe Trp Ile His
            235                 240                 245

cag gtc acc tgc ctg ggg aca gag ccc cac atg gcc aac tgc cag gtg   883
Gln Val Thr Cys Leu Gly Thr Glu Pro His Met Ala Asn Cys Gln Val
        250                 255                 260

cag gtg gct cca gcc cgg ggc aag ctg cgg cca gcc tgc cca ggt ggc   931
Gln Val Ala Pro Ala Arg Gly Lys Leu Arg Pro Ala Cys Pro Gly Gly
    265                 270                 275

atg cac gct gtg gtc agc tgt gtg gca ggg cct cac ttc cgc cca ccg   979
Met His Ala Val Val Ser Cys Val Ala Gly Pro His Phe Arg Pro Pro
280                 285                 290                 295

aag aca aag cca caa cgc aaa ggg tcc tgg gca gag gag ccg agg gtg   1027
Lys Thr Lys Pro Gln Arg Lys Gly Ser Trp Ala Glu Glu Pro Arg Val
                300                 305                 310

cgc ctg cgc tcc ggg gcc cag gtg ggc gag ggc cgg gtg gaa gtg ctc   1075
Arg Leu Arg Ser Gly Ala Gln Val Gly Glu Gly Arg Val Glu Val Leu
            315                 320                 325

atg aac cgc cag tgg ggc acg gtc tgt gac cac agg tgg aac ctc atc   1123
Met Asn Arg Gln Trp Gly Thr Val Cys Asp His Arg Trp Asn Leu Ile
        330                 335                 340

tct gcc agt gtc gtg tgt cgt cag ctg ggc ttt ggc tct gct cgg gag   1171
Ser Ala Ser Val Val Cys Arg Gln Leu Gly Phe Gly Ser Ala Arg Glu
```

```
    345                 350                 355

gcc ctc ttt ggg gcc cgg ctg ggc caa ggg cta ggg ccc atc cac ctg   1219
Ala Leu Phe Gly Ala Arg Leu Gly Gln Gly Leu Gly Pro Ile His Leu
360                 365                 370                 375

agt gag gtg cgc tgc agg gga tat gag cgg acc ctc agc gac tgc cct   1267
Ser Glu Val Arg Cys Arg Gly Tyr Glu Arg Thr Leu Ser Asp Cys Pro
                380                 385                 390

gcc ctg gaa ggg tcc cag aat ggt tgc caa cat gag aat gat gct gct   1315
Ala Leu Glu Gly Ser Gln Asn Gly Cys Gln His Glu Asn Asp Ala Ala
            395                 400                 405

gtc agg tgc aat gtc cct aac atg ggc ttt cag aat cag gtg cgc ttg   1363
Val Arg Cys Asn Val Pro Asn Met Gly Phe Gln Asn Gln Val Arg Leu
        410                 415                 420

gct ggt ggg cgt atc cct gag gag ggg cta ttg gag gtg cag gtg gag   1411
Ala Gly Gly Arg Ile Pro Glu Glu Gly Leu Leu Glu Val Gln Val Glu
    425                 430                 435

gtg aac ggg gtc cca cgc tgg ggg agc gtg tgc agt gaa aac tgg ggg   1459
Val Asn Gly Val Pro Arg Trp Gly Ser Val Cys Ser Glu Asn Trp Gly
440                 445                 450                 455

ctc acc gaa gcc atg gtg gcc tgc cga cag ctc ggc ctg ggt ttt gcc   1507
Leu Thr Glu Ala Met Val Ala Cys Arg Gln Leu Gly Leu Gly Phe Ala
                460                 465                 470

atc cat gcc tac aag gaa acc tgg ttc tgg tcg ggg acg cca agg gcc   1555
Ile His Ala Tyr Lys Glu Thr Trp Phe Trp Ser Gly Thr Pro Arg Ala
            475                 480                 485

cag gag gtg gtg atg agt ggg gtg cgc tgc tca ggc aca gag ctg gcc   1603
Gln Glu Val Val Met Ser Gly Val Arg Cys Ser Gly Thr Glu Leu Ala
        490                 495                 500

ctg cag cag tgc cag agg cac ggg ccg gtg cac tgc tcc cac ggt ggc   1651
Leu Gln Gln Cys Gln Arg His Gly Pro Val His Cys Ser His Gly Gly
    505                 510                 515

ggg cgc ttc ctg gct gga gtc tcc tgc atg gac agt gca cca gac ctg   1699
Gly Arg Phe Leu Ala Gly Val Ser Cys Met Asp Ser Ala Pro Asp Leu
520                 525                 530                 535

gtg atg aac gcc cag cta gtg cag gag acg gcc tac ttg gag gac cgc   1747
Val Met Asn Ala Gln Leu Val Gln Glu Thr Ala Tyr Leu Glu Asp Arg
                540                 545                 550

ccg ctc agc cag ctg tat tgt gcc cac gag gag aac tgc ctc tcc aag   1795
Pro Leu Ser Gln Leu Tyr Cys Ala His Glu Glu Asn Cys Leu Ser Lys
            555                 560                 565

tct gca gat cac atg gac tgg ccc tac gga tac cgc cgc cta ttg cgc   1843
Ser Ala Asp His Met Asp Trp Pro Tyr Gly Tyr Arg Arg Leu Leu Arg
        570                 575                 580

ttc tcc aca cag atc tac aat ctg ggc cgg act gac ttt cgt cca aag   1891
Phe Ser Thr Gln Ile Tyr Asn Leu Gly Arg Thr Asp Phe Arg Pro Lys
    585                 590                 595
```

```
act gga cgc gat agc tgg gtt tgg cac cag tgc cac agg cat tac cac   1939
Thr Gly Arg Asp Ser Trp Val Trp His Gln Cys His Arg His Tyr His
600                 605                 610                 615

agc att gag gtc ttc acc cac tac gac ctc ctc act ctc aat ggc tcc   1987
Ser Ile Glu Val Phe Thr His Tyr Asp Leu Leu Thr Leu Asn Gly Ser
                620                 625                 630

aag gtg gct gag ggg cac aag gcc agc ttc tgt ctg gag gac aca aac   2035
Lys Val Ala Glu Gly His Lys Ala Ser Phe Cys Leu Glu Asp Thr Asn
            635                 640                 645

tgc ccc aca gga ctg cag cgg cgc tac gca tgt gcc aac ttt gga gaa   2083
Cys Pro Thr Gly Leu Gln Arg Arg Tyr Ala Cys Ala Asn Phe Gly Glu
        650                 655                 660

cag gga gtg act gta ggc tgc tgg gac acc tac cgg cat gac att gat   2131
Gln Gly Val Thr Val Gly Cys Trp Asp Thr Tyr Arg His Asp Ile Asp
    665                 670                 675

tgc cag tgg gtg gat atc aca gat gtg ggc ccc ggg aat tat atc ttc   2179
Cys Gln Trp Val Asp Ile Thr Asp Val Gly Pro Gly Asn Tyr Ile Phe
680                 685                 690                 695

cag gtg att gtg aac ccc cac tat gaa gtg gca gag tca gat ttc tcc   2227
Gln Val Ile Val Asn Pro His Tyr Glu Val Ala Glu Ser Asp Phe Ser
                700                 705                 710

aac aat atg ctg cag tgc cgc tgc aag tat gat ggg cac cgg gtc tgg   2275
Asn Asn Met Leu Gln Cys Arg Cys Lys Tyr Asp Gly His Arg Val Trp
            715                 720                 725

ctg cac aac tgc cac aca ggg aat tca tac cca gcc aat gca gaa ctc   2323
Leu His Asn Cys His Thr Gly Asn Ser Tyr Pro Ala Asn Ala Glu Leu
        730                 735                 740

tcc ctg gag cag gaa cag cgt ctc agg aac aac ctc atc tgaagctgtc    2372
Ser Leu Glu Gln Glu Gln Arg Leu Arg Asn Asn Leu Ile
    745                 750                 755

actgcacact cctagctgct gccgatacac cagatacctc agcttattgg agccatgccc 2432
ttcacagagt cccaactcag aggaaaaggg ccagtgccaa ggggcaccaa gaacctgctc 2492
aggaagcctt ttgatggcaa gatcaccaat ccagatggta ttgctccctc aggatggctc 2552
tgggcctgcc cctaagggcc tgtggcctat ggaatatgtc ctccaggctt tgcttagctg 2612
agctcctctt ctgtaaggaa acccagtcat ccctgaatct tgccacagag atccgggatt 2672
caggagctct cagtttctta gggatggact atggcccagt cccccatcta agtggtgctt 2732
tgcaaatgtc ttggaggagt ataggacaga ggaccaaaat acacagcagg tagtgktagc 2792
tctctgctag gagctcaaag caacacaact tgtatcaaaa tcacaactgg cagaaaaact 2852
gggtggatcc aatcctttct ttcatctgtg tatttaagaa yyaaccytta mactctggtc 2912
tttaggggcc ttamctttat ttwmcacama maatnggngg ttttattatc yttggaagca 2972
caaa                                                              2976

<210> 2
<211> 756
<212> PRT
<213> Homo sapiens

<400> 2
Met Ala Trp Ser Pro Pro Ala Thr Leu Phe Leu Phe Leu Leu Leu Leu
 1               5                  10                  15
```

```
Gly Gln Pro Pro Pro Ser Arg Pro Gln Ser Leu Gly Thr Thr Lys Leu
            20                  25                  30
Arg Leu Val Gly Pro Glu Ser Lys Pro Glu Glu Gly Arg Leu Glu Val
        35                  40                  45
Leu His Gln Gly Gln Trp Gly Thr Val Cys Asp Asp Asn Phe Ala Ile
    50                  55                  60
Gln Glu Ala Thr Val Ala Cys Arg Gln Leu Gly Phe Glu Ala Ala Leu
65                  70                  75                  80
Thr Trp Ala His Ser Ala Lys Tyr Gly Gln Gly Glu Gly Pro Ile Trp
                85                  90                  95
Leu Asp Asn Val Arg Cys Val Gly Thr Glu Ser Ser Leu Asp Gln Cys
            100                 105                 110
Gly Ser Asn Gly Trp Gly Val Ser Asp Cys Ser His Ser Glu Asp Val
        115                 120                 125
Gly Val Ile Cys His Pro Arg Arg His Arg Gly Tyr Leu Ser Glu Thr
    130                 135                 140
Val Ser Asn Ala Leu Gly Pro Gln Gly Arg Arg Leu Glu Glu Val Arg
145                 150                 155                 160
Leu Lys Pro Ile Leu Ala Ser Ala Lys Gln His Ser Pro Val Thr Glu
                165                 170                 175
Gly Ala Val Glu Val Lys Tyr Glu Gly His Trp Arg Gln Val Cys Asp
            180                 185                 190
Gln Gly Trp Thr Met Asn Asn Ser Arg Val Val Cys Gly Met Leu Gly
        195                 200                 205
Phe Pro Ser Glu Val Pro Val Asp Ser His Tyr Tyr Arg Lys Val Trp
    210                 215                 220
Asp Leu Lys Met Arg Asp Pro Lys Ser Arg Leu Lys Ser Leu Thr Asn
225                 230                 235                 240
Lys Asn Ser Phe Trp Ile His Gln Val Thr Cys Leu Gly Thr Glu Pro
                245                 250                 255
His Met Ala Asn Cys Gln Val Gln Val Ala Pro Ala Arg Gly Lys Leu
            260                 265                 270
Arg Pro Ala Cys Pro Gly Gly Met His Ala Val Val Ser Cys Val Ala
        275                 280                 285
Gly Pro His Phe Arg Pro Pro Lys Thr Lys Pro Gln Arg Lys Gly Ser
    290                 295                 300
Trp Ala Glu Glu Pro Arg Val Arg Leu Arg Ser Gly Ala Gln Val Gly
305                 310                 315                 320
Glu Gly Arg Val Glu Val Leu Met Asn Arg Gln Trp Gly Thr Val Cys
                325                 330                 335
Asp His Arg Trp Asn Leu Ile Ser Ala Ser Val Val Cys Arg Gln Leu
            340                 345                 350
Gly Phe Gly Ser Ala Arg Glu Ala Leu Phe Gly Ala Arg Leu Gly Gln
        355                 360                 365
Gly Leu Gly Pro Ile His Leu Ser Glu Val Arg Cys Arg Gly Tyr Glu
    370                 375                 380
Arg Thr Leu Ser Asp Cys Pro Ala Leu Glu Gly Ser Gln Asn Gly Cys
385                 390                 395                 400
Gln His Glu Asn Asp Ala Ala Val Arg Cys Asn Val Pro Asn Met Gly
                405                 410                 415
Phe Gln Asn Gln Val Arg Leu Ala Gly Gly Arg Ile Pro Glu Glu Gly
            420                 425                 430
Leu Leu Glu Val Gln Val Glu Val Asn Gly Val Pro Arg Trp Gly Ser
        435                 440                 445
Val Cys Ser Glu Asn Trp Gly Leu Thr Glu Ala Met Val Ala Cys Arg
    450                 455                 460
Gln Leu Gly Leu Gly Phe Ala Ile His Ala Tyr Lys Glu Thr Trp Phe
465                 470                 475                 480
Trp Ser Gly Thr Pro Arg Ala Gln Glu Val Val Met Ser Gly Val Arg
                485                 490                 495
Cys Ser Gly Thr Glu Leu Ala Leu Gln Gln Cys Gln Arg His Gly Pro
```

```
            500                 505                 510
Val His Cys Ser His Gly Gly Gly Arg Phe Leu Ala Gly Val Ser Cys
        515                 520                 525
Met Asp Ser Ala Pro Asp Leu Val Met Asn Ala Gln Leu Val Gln Glu
    530                 535                 540
Thr Ala Tyr Leu Glu Asp Arg Pro Leu Ser Gln Leu Tyr Cys Ala His
545                 550                 555                 560
Glu Glu Asn Cys Leu Ser Lys Ser Ala Asp His Met Asp Trp Pro Tyr
                565                 570                 575
Gly Tyr Arg Arg Leu Leu Arg Phe Ser Thr Gln Ile Tyr Asn Leu Gly
            580                 585                 590
Arg Thr Asp Phe Arg Pro Lys Thr Gly Arg Asp Ser Trp Val Trp His
        595                 600                 605
Gln Cys His Arg His Tyr His Ser Ile Glu Val Phe Thr His Tyr Asp
    610                 615                 620
Leu Leu Thr Leu Asn Gly Ser Lys Val Ala Glu Gly His Lys Ala Ser
625                 630                 635                 640
Phe Cys Leu Glu Asp Thr Asn Cys Pro Thr Gly Leu Gln Arg Arg Tyr
                645                 650                 655
Ala Cys Ala Asn Phe Gly Glu Gln Gly Val Thr Val Gly Cys Trp Asp
            660                 665                 670
Thr Tyr Arg His Asp Ile Asp Cys Gln Trp Val Asp Ile Thr Asp Val
        675                 680                 685
Gly Pro Gly Asn Tyr Ile Phe Gln Val Ile Val Asn Pro His Tyr Glu
    690                 695                 700
Val Ala Glu Ser Asp Phe Ser Asn Asn Met Leu Gln Cys Arg Cys Lys
705                 710                 715                 720
Tyr Asp Gly His Arg Val Trp Leu His Asn Cys His Thr Gly Asn Ser
                725                 730                 735
Tyr Pro Ala Asn Ala Glu Leu Ser Leu Glu Gln Glu Gln Arg Leu Arg
            740                 745                 750
Asn Asn Leu Ile
        755
```

<210> 3
<211> 2268
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)...(2268)

<400> 3

```
atg gcg tgg tcc cca cca gcc acc ctc ttt ctg ttc ctg ctg ctg cta   48
Met Ala Trp Ser Pro Pro Ala Thr Leu Phe Leu Phe Leu Leu Leu Leu
 1               5                  10                  15

ggc cag ccc cct ccc agc agg cca cag tca ctg ggc acc act aag ctc   96
Gly Gln Pro Pro Pro Ser Arg Pro Gln Ser Leu Gly Thr Thr Lys Leu
            20                  25                  30

cgg ctg gtg ggc cca gag agc aag cca gag gag ggc cgc ctg gag gtg   144
Arg Leu Val Gly Pro Glu Ser Lys Pro Glu Glu Gly Arg Leu Glu Val
        35                  40                  45

ctg cac cag ggc cag tgg ggc acc gtg tgt gat gac aac ttt gct atc   192
Leu His Gln Gly Gln Trp Gly Thr Val Cys Asp Asp Asn Phe Ala Ile
    50                  55                  60
```

```
cag gag gcc aca gtg gct tgc cgc cag ctg ggc ttc gaa gct gcc ttg   240
Gln Glu Ala Thr Val Ala Cys Arg Gln Leu Gly Phe Glu Ala Ala Leu
 65                  70                  75                  80

acc tgg gcc cac agt gcc aag tac ggc caa ggg gag gga ccc atc tgg   288
Thr Trp Ala His Ser Ala Lys Tyr Gly Gln Gly Glu Gly Pro Ile Trp
                 85                  90                  95

ctg gac aat gtg cgc tgt gtg ggc aca gag agc tcc ttg gac cag tgc   336
Leu Asp Asn Val Arg Cys Val Gly Thr Glu Ser Ser Leu Asp Gln Cys
            100                 105                 110

ggg tct aat ggc tgg gga gtc agt gac tgc agt cac tca gaa gac gta   384
Gly Ser Asn Gly Trp Gly Val Ser Asp Cys Ser His Ser Glu Asp Val
        115                 120                 125

ggg gtg ata tgc cac ccc cgg cgc cat cgt ggc tac ctt tct gaa act   432
Gly Val Ile Cys His Pro Arg Arg His Arg Gly Tyr Leu Ser Glu Thr
    130                 135                 140

gtc tcc aat gcc ctt ggg ccc cag ggc cgg cgg ctg gag gag gtg cgg   480
Val Ser Asn Ala Leu Gly Pro Gln Gly Arg Arg Leu Glu Glu Val Arg
145                 150                 155                 160

ctc aag ccc atc ctt gcc agt gcc aag cag cat agc cca gtg acc gag   528
Leu Lys Pro Ile Leu Ala Ser Ala Lys Gln His Ser Pro Val Thr Glu
                165                 170                 175

gga gcc gtg gag gtg aag tat gag ggc cac tgg cgg cag gtg tgt gac   576
Gly Ala Val Glu Val Lys Tyr Glu Gly His Trp Arg Gln Val Cys Asp
            180                 185                 190

cag ggc tgg acc atg aac aac agc agg gtg gtg tgc ggg atg ctg ggc   624
Gln Gly Trp Thr Met Asn Asn Ser Arg Val Val Cys Gly Met Leu Gly
        195                 200                 205

ttc ccc agc gag gtg cct gtc gac agc cac tac tac agg aaa gtc tgg   672
Phe Pro Ser Glu Val Pro Val Asp Ser His Tyr Tyr Arg Lys Val Trp
    210                 215                 220

gat ctg aag atg agg gac cct aag tct agg ctg aag agc ctg acg aat   720
Asp Leu Lys Met Arg Asp Pro Lys Ser Arg Leu Lys Ser Leu Thr Asn
225                 230                 235                 240

aag aac tcc ttc tgg atc cac cag gtc acc tgc ctg ggg aca gag ccc   768
Lys Asn Ser Phe Trp Ile His Gln Val Thr Cys Leu Gly Thr Glu Pro
                245                 250                 255

cac atg gcc aac tgc cag gtg cag gtg gct cca gcc cgg ggc aag ctg   816
His Met Ala Asn Cys Gln Val Gln Val Ala Pro Ala Arg Gly Lys Leu
            260                 265                 270

cgg cca gcc tgc cca ggt ggc atg cac gct gtg gtc agc tgt gtg gca   864
Arg Pro Ala Cys Pro Gly Gly Met His Ala Val Val Ser Cys Val Ala
        275                 280                 285

ggg cct cac ttc cgc cca ccg aag aca aag cca caa cgc aaa ggg tcc   912
Gly Pro His Phe Arg Pro Pro Lys Thr Lys Pro Gln Arg Lys Gly Ser
    290                 295                 300

tgg gca gag gag ccg agg gtg cgc ctg cgc tcc ggg gcc cag gtg ggc   960
```

```
Trp Ala Glu Glu Pro Arg Val Arg Leu Arg Ser Gly Ala Gln Val Gly
305                 310                 315                 320

gag ggc cgg gtg gaa gtg ctc atg aac cgc cag tgg ggc acg gtc tgt   1008
Glu Gly Arg Val Glu Val Leu Met Asn Arg Gln Trp Gly Thr Val Cys
                325                 330                 335

gac cac agg tgg aac ctc atc tct gcc agt gtc gtg tgt cgt cag ctg   1056
Asp His Arg Trp Asn Leu Ile Ser Ala Ser Val Val Cys Arg Gln Leu
            340                 345                 350

ggc ttt ggc tct gct cgg gag gcc ctc ttt ggg gcc cgg ctg ggc caa   1104
Gly Phe Gly Ser Ala Arg Glu Ala Leu Phe Gly Ala Arg Leu Gly Gln
        355                 360                 365

ggg cta ggg ccc atc cac ctg agt gag gtg cgc tgc agg gga tat gag   1152
Gly Leu Gly Pro Ile His Leu Ser Glu Val Arg Cys Arg Gly Tyr Glu
    370                 375                 380

cgg acc ctc agc gac tgc cct gcc ctg gaa ggg tcc cag aat ggt tgc   1200
Arg Thr Leu Ser Asp Cys Pro Ala Leu Glu Gly Ser Gln Asn Gly Cys
385                 390                 395                 400

caa cat gag aat gat gct gct gtc agg tgc aat gtc cct aac atg ggc   1248
Gln His Glu Asn Asp Ala Ala Val Arg Cys Asn Val Pro Asn Met Gly
                405                 410                 415

ttt cag aat cag gtg cgc ttg gct ggt ggg cgt atc cct gag gag ggg   1296
Phe Gln Asn Gln Val Arg Leu Ala Gly Gly Arg Ile Pro Glu Glu Gly
            420                 425                 430

cta ttg gag gtg cag gtg gag gtg aac ggg gtc cca cgc tgg ggg agc   1344
Leu Leu Glu Val Gln Val Glu Val Asn Gly Val Pro Arg Trp Gly Ser
        435                 440                 445

gtg tgc agt gaa aac tgg ggg ctc acc gaa gcc atg gtg gcc tgc cga   1392
Val Cys Ser Glu Asn Trp Gly Leu Thr Glu Ala Met Val Ala Cys Arg
    450                 455                 460

cag ctc ggc ctg ggt ttt gcc atc cat gcc tac aag gaa acc tgg ttc   1440
Gln Leu Gly Leu Gly Phe Ala Ile His Ala Tyr Lys Glu Thr Trp Phe
465                 470                 475                 480

tgg tcg ggg acg cca agg gcc cag gag gtg gtg atg agt ggg gtg cgc   1488
Trp Ser Gly Thr Pro Arg Ala Gln Glu Val Val Met Ser Gly Val Arg
                485                 490                 495

tgc tca ggc aca gag ctg gcc ctg cag cag tgc cag agg cac ggg ccg   1536
Cys Ser Gly Thr Glu Leu Ala Leu Gln Gln Cys Gln Arg His Gly Pro
            500                 505                 510

gtg cac tgc tcc cac ggt ggc ggg cgc ttc ctg gct gga gtc tcc tgc   1584
Val His Cys Ser His Gly Gly Gly Arg Phe Leu Ala Gly Val Ser Cys
        515                 520                 525

atg gac agt gca cca gac ctg gtg atg aac gcc cag cta gtg cag gag   1632
Met Asp Ser Ala Pro Asp Leu Val Met Asn Ala Gln Leu Val Gln Glu
    530                 535                 540

acg gcc tac ttg gag gac cgc ccg ctc agc cag ctg tat tgt gcc cac   1680
Thr Ala Tyr Leu Glu Asp Arg Pro Leu Ser Gln Leu Tyr Cys Ala His
```

```
545                 550                 555                 560

gag gag aac tgc ctc tcc aag tct gca gat cac atg gac tgg ccc tac   1728
Glu Glu Asn Cys Leu Ser Lys Ser Ala Asp His Met Asp Trp Pro Tyr
                565                 570                 575

gga tac cgc cgc cta ttg cgc ttc tcc aca cag atc tac aat ctg ggc   1776
Gly Tyr Arg Arg Leu Leu Arg Phe Ser Thr Gln Ile Tyr Asn Leu Gly
            580                 585                 590

cgg act gac ttt cgt cca aag act gga cgc gat agc tgg gtt tgg cac   1824
Arg Thr Asp Phe Arg Pro Lys Thr Gly Arg Asp Ser Trp Val Trp His
        595                 600                 605

cag tgc cac agg cat tac cac agc att gag gtc ttc acc cac tac gac   1872
Gln Cys His Arg His Tyr His Ser Ile Glu Val Phe Thr His Tyr Asp
    610                 615                 620

ctc ctc act ctc aat ggc tcc aag gtg gct gag ggg cac aag gcc agc   1920
Leu Leu Thr Leu Asn Gly Ser Lys Val Ala Glu Gly His Lys Ala Ser
625                 630                 635                 640

ttc tgt ctg gag gac aca aac tgc ccc aca gga ctg cag cgg cgc tac   1968
Phe Cys Leu Glu Asp Thr Asn Cys Pro Thr Gly Leu Gln Arg Arg Tyr
                645                 650                 655

gca tgt gcc aac ttt gga gaa cag gga gtg act gta ggc tgc tgg gac   2016
Ala Cys Ala Asn Phe Gly Glu Gln Gly Val Thr Val Gly Cys Trp Asp
            660                 665                 670

acc tac cgg cat gac att gat tgc cag tgg gtg gat atc aca gat gtg   2064
Thr Tyr Arg His Asp Ile Asp Cys Gln Trp Val Asp Ile Thr Asp Val
        675                 680                 685

ggc ccc ggg aat tat atc ttc cag gtg att gtg aac ccc cac tat gaa   2112
Gly Pro Gly Asn Tyr Ile Phe Gln Val Ile Val Asn Pro His Tyr Glu
    690                 695                 700

gtg gca gag tca gat ttc tcc aac aat atg ctg cag tgc cgc tgc aag   2160
Val Ala Glu Ser Asp Phe Ser Asn Asn Met Leu Gln Cys Arg Cys Lys
705                 710                 715                 720

tat gat ggg cac cgg gtc tgg ctg cac aac tgc cac aca ggg aat tca   2208
Tyr Asp Gly His Arg Val Trp Leu His Asn Cys His Thr Gly Asn Ser
                725                 730                 735

tac cca gcc aat gca gaa ctc tcc ctg gag cag gaa cag cgt ctc agg   2256
Tyr Pro Ala Asn Ala Glu Leu Ser Leu Glu Gln Glu Gln Arg Leu Arg
            740                 745                 750

aac aac ctc atc                                                 2268
Asn Asn Leu Ile
        755
```

<210> 4
<211> 9
<212> PRT
<213> Homo sapiens

<220>

<221> VARIANT
<222> (2)...(2)
<223> Xaa= any amino acid

<221> VARIANT
<222> (5)...(5)
<223> Xaa= any amino acid

<400> 4
Trp Xaa Trp His Xaa Cys His Tyr His
1 5

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 December 2001 (06.12.2001)

PCT

(10) International Publication Number
WO 01/92495 A3

(51) International Patent Classification⁷: C12N 15/53, 9/06, C07K 16/40, G01N 33/573, C12Q 1/68

(21) International Application Number: PCT/US01/17405

(22) International Filing Date: 29 May 2001 (29.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/207,650 26 May 2000 (26.05.2000) US

(71) Applicant *(for all designated States except US)*: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: MEYERS, Rachel [US/US]; 115 Devonshire Road, Newton, MA 02468 (US).

(74) Agents: MANDRAGOURAS, Amy, E.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
with international search report
before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
2 May 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/92495 A3

(54) Title: A HUMAN LYSYL IXODASE (47765) AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acid molecules, designated LSO nucleic acid molecules, which encode novel LSO-related lysyl oxidase molecules. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing LSO nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which an LSO gene has been introduced or disrupted. The invention still further provides isolated LSO proteins, fusion proteins, antigenic peptides and anti-LSO antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

I**TERNATIONAL SEARCH REPORT

I. ational Application No
PCT/US 01/17405

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/53 C12N9/06 C07K16/40 G01N33/573 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBL, WPI Data, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! 17 March 1997 (1997-03-17) "Human lysyl oxidase-related protein (WS9-14) mRNA" Database accession no. U89942 XP002186675 the whole document	5-26
Y	-& SAITO H ET AL.: "Regulation of a novel gene encoding a lysyl oxidase-related protein in cellular adhesion and senescence." J. BIOL. CHEM, vol. 272, no. 13, 28 March 1997 (1997-03-28), pages 8157-8160, XP002187526 the whole document	1-26

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 January 2002	26/02/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Bucka, A

6

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
Y	WO 98 06830 A (UNIV SYDNEY ;WEISS ANTHONY STEVEN (AU)) 19 February 1998 (1998-02-19) the whole document	1-26
P,X	WO 00 61774 A (HUMAN GENOME SCIENCES INC; NI JIAN (US); ROSEN CRAIG A (US); RUBEN) 19 October 2000 (2000-10-19) SEQ ID: 3, 7, 9 & 24	5-26
P,X	WO 00 58463 A (KUMBLE KRISHANAND D; ONRUST RENE (NZ); ABERNETHY NEVIN (NZ); SLEEM) 5 October 2000 (2000-10-05) SEQ ID: 44, 52	5-26
P,X	WO 00 44910 A (KHODADOUST MEHRAN M; MACBETH KYLE J (US); MILLENNIUM PHARM INC (US) 3 August 2000 (2000-08-03) SEQ ID: 1 & 3	5-26
E	WO 01 79291 A (BRUNS CHRISTOPHER M ;INCYTE GENOMICS INC (US); PATTERSON CHANDRA () 25 October 2001 (2001-10-25) SEQ ID: 17	5-26
X	WO 00 29422 A (FLORENCE KIMBERLY A ;HUMAN GENOME SCIENCES INC (US); NI JIAN (US);) 25 May 2000 (2000-05-25) SEQ ID: 17 & 43	5-26
X	DATABASE EMBL 'Online! 3 June 1999 (1999-06-03) "Homo sapiens lysyl oxidase-like protein 2 (LOXL2) mRNA" Database accession no. AF117949 XP002186676 the whole document -& SAUX CJ ET AL.: "The LOXL2 gene encodes a new lysyl oxidase-like protein and is expressed at high levels in reproductive tissues" J. BIOL. CHEM., vol. 274, no. 18, 30 April 1999 (1999-04-30), pages 12939-12944, XP002186674 the whole document	5-26
X	DATABASE EMBL 'Online! 29 June 1999 (1999-06-29) "cn10c08.y1 Normal human Trabecular Bone Cells Homo sapiens cDNA clone NHTBC_cn10c08 random, mRNA sequence" Database accession no. AI751493 XP002186678 the whole document	5-16

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claim 3 is incomprehensible, because it refers to a plasmid deposited with the ATCC, for which the accession number is not provided. Consequently, claim 3 does not meet the requirements of Article 6 PCT, because it does not define the matter for which protection is sought. A meaningful search of the claim is therefore impossible.

Present claims 17, 19, 22 and 25 at least partly relate to compounds defined by reference to a desirable characteristic or property, namely their ability to bind to the claimed proteins (claims 17, 19, 25) or nucleic acids (claim 22), or to modulate the activity of the polypeptide by binding to it (claim 25).
The claims cover all compounds having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and disclosure within the meaning of Article 5 PCT for only a very limited number of such compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the compounds by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to antibodies, as mentioned in claim 18.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/17405

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9806830	A	19-02-1998	AU	733946 B2	31-05-2001
			AU	3689497 A	06-03-1998
			WO	9806830 A1	19-02-1998
			EP	0960192 A1	01-12-1999
			JP	2000517172 T	26-12-2000
WO 0061774	A	19-10-2000	AU	4073600 A	14-11-2000
			AU	4081000 A	14-11-2000
			AU	4451300 A	14-11-2000
			WO	0061625 A1	19-10-2000
			WO	0061774 A2	19-10-2000
			WO	0061614 A2	19-10-2000
			AU	5457000 A	02-01-2001
			WO	0077021 A1	21-12-2000
WO 0058463	A	05-10-2000	US	6242419 B1	05-06-2001
			AU	2702100 A	16-10-2000
			EP	1163334 A1	19-12-2001
			WO	0058463 A1	05-10-2000
WO 0044910	A	03-08-2000	US	6140056 A	31-10-2000
			US	6300092 B1	09-10-2001
			AU	3217100 A	18-08-2000
			EP	1149169 A1	31-10-2001
			WO	0044910 A1	03-08-2000
WO 0179291	A	25-10-2001	WO	0179291 A2	25-10-2001
WO 0029422	A	25-05-2000	AU	1716200 A	05-06-2000
			EP	1137656 A1	04-10-2001
			WO	0029422 A1	25-05-2000